Oral Tradition, 8.1 (1993): 87-142

Homer and the *Roland*: The Shared Formular Technique, Part I

William Merritt Sale

I. Introduction

The argument of the following article, though necessarily long and demanding, can be summarized briefly. Homer employs his noun-formulae consistently, so that the principles of their employment can be stated mathematically in the form of equations and graphs. So too does the poet of the *Chanson de Roland*. Moreover, each displays virtually the same equations as the other: in the employment of formulae, the techniques of the two are almost identical. The similarity is particularly arresting when we observe that it results from the pervasive use of infrequent formulae, formulae that occur very often, but only a few times each. A great many of these infrequent formulae are either combinations of nouns with standardized adjectives or verbs (called "generic words"), or are flexible formulae, phrases that can be separated, inverted, inflected, or moved about in the line. Such adjectives and such formulae are equipment intended to meet poetic needs that arise very commonly as a type, but individually very rarely; and it is very hard to avoid the inference that these are needs that arise in the course of composing poetry during an oral performance. Even from Homer alone, or from the *Roland* poet alone, we could infer a technique of oral composition. We then ask why their mathematically analyzable compositional principles should be so very similar, noticing meanwhile that these principles are also shared by the *Cantar de Mio Cid* and the twentieth-century Yugoslavian oral poet Avdo Međedović, but not by Apollonius of Rhodes, Virgil, or Quintus of Smyrna, though the latter especially is a highly formulaic poet. If a technique almost certainly designed to meet the exigencies of oral composition in performance is shared by four poets believed on other grounds (or, in the case of Avdo, known) to be an oral poet, and is eschewed by three imitators of Homer known to have written, we are approaching scientific demonstration that Homer and these medieval poets composed orally. Many readers of *Oral Tradition* may feel that we knew this already from the work of Milman Parry, Albert Lord, Joseph Duggan, and others. But among classicists, at any rate, Parry's position has recently become somewhat beleaguered; it is important to see how close we are to proving the truth of Lord's opinion, that Homer did what Avdo did and dictated a text to a scribe—or at least did something very much like it.

Since the following comparison of the styles of the *Iliad*, the *Odyssey*, and the *Chanson de Roland* is based upon statistics, it may be appropriate to say a word or two at the beginning on how statistics has previously been applied to the three poems. All three are long, and all three are repetitious, and these two facts combine to put a useful scholarly tool into our hands: we can count the number of occurrences of phenomena important to the style and arrive at numbers that are statistically significant, numbers large enough to enable meaningful comparisons. Milman Parry counted the number of occurrences of formulae in order to show, for instance, how large a number of different formulae belong to a given system (that is, possess the same meter and syntax; Parry 1971:17). He counted occurrences of nouns with and without an epithet, in order to compare the behavior of nouns that had different metrical shapes, and to compare nouns in Homer with nouns in Virgil (34-36). Denys Page counted words and phrases that occur frequently in the *Odyssey* but are absent or very rare in the *Iliad* (and vice versa) in an effort to show that the poems were composed in different geographical places (1955:149-55). Eugene O'Neill counted and compared the number of times metrical word-types occur in various locations in the hexameter line (1942). O'Neill went beyond merely counting and comparing by stating the percentage of times a word-type appeared in a given position, revealing thereby the commonest—the favored—positions. Albert Lord made calculations of formulaic density (the percentage of lines or half-lines in a given sample that are formulaic) for Yugoslavian, Homeric, Old English, and other poetry (1960). Joseph Duggan took a further step by calculating the formulaic density of entire poems: the *Chanson de Roland*, the *Cantar de Mio Cid*, and other *chansons de geste* (1973, 1975:74-83). He showed that certain poems, including the *Roland* and the Cid, have a much higher formulaic density than others, a fact that allowed him to argue that high formulaic density must be due to the fact that the poems were orally composed. Margalit Finkelberg counted and compared occurrences of verb-formulae to determine their formularity, the percentage of their formulaic occurrences out of all their occurrences (1989:179-87). I myself counted numbers of occurrences of Homeric formulae for *Olympos* and *Ouranos* meaning "the divine home," in order to show that the set of *Olympos* formulae was the earlier (1984); I then counted place-phrases in the *Iliad* and calculated the percentage of their formulaic occurrences out of all their occurrences (their **formularity**) in order to expose a remarkable deficiency in formulae meaning "in Troy" and "from Troy" (1987); and I counted the number of occurrences of all nominative proper-noun formulae in order to show that the Trojans lacked regular formulae (formulae exactly repeated 6 times or more) (1989).

In the 1987 article, with the very considerable help of Professor Dee Clayman of CUNY, I used a statistical test to prove the significance of the deficiency in the Trojan place-phrases; and in the 1989 piece I employed the same test to evaluate the uniformity of the formularity of the nouns I was studying. In this last article I also developed an *equation* to plot the relationship between localization (the percentage of times a word occurs in that place in the hexameter line in which it occurs most often) and regularity (the percentage of occurrences of regular formulae out of all formulaic occurrences). Meanwhile, Richard Janko had used statistical tests to show the significance of linguistic changes for evaluating the relative time of composition of the Homeric and Hesiodic poems and the Homeric Hymns (1982): the greater the extent of change, the later the poem was likely to be.

The current article uses simple percentages and equations to compare the formularity of nouns in Homer and the *Roland*. But then it deepens the scope of the statistical study of epic verse by using equations and graphs to get at more subtle aspects of formulaic composition: the number of different formulae a noun displays relative to the total number of its occurrences, and the difference in behavior between a poem's frequently- and infrequently-occurring formulae. The *equations* are linear, but more complex than any I have constructed hitherto, since they entail as many as four variables. The *graph*, however, is no longer linear, but hyperbolic—a further complexity, but not one that steps up the demands upon our mathematical experience, since it is not the equation that interests us in this article, but the graph itself. The technical sophistication required by these ideas therefore falls well within the scope of basic algebra and the simple statistics of fitting a linear curve to plottable data (and indeed we use a computer program to determine the best fit!). The real difficulty offered by what follows is not mathematical; it comes not from the curves, but from their explanation. Four variables acting together in a linear equation are easy enough to handle algebraically; what is harder is to picture the activity the equation symbolizes, especially since this is the activity not of the poet, but of his nouns. And the mathematics of the hyperbola is not relevant to the current study; we want to know what formulaic behavior engendered such a shape. Hence the energy of the following description is directed primarily at explaining the phenomena, such as **different formulae** and **formulaic occurrences**, to which the equations refer. (These phrases will be highlighted with bold font when when they refer to variables, that is, when the mathematical quality or behavior of their referents is being stressed.)

There are three arresting statistical correspondences between the *Iliad* and *Odyssey* of Homer and the Oxford *Roland*.[1] These correspondences are almost certainly due to similarities in the compositional techniques of the three poems, but in much of the ensuing discussion it will prove easier to think of them as due to the behavior of the nouns themselves. In speaking thus metaphorically of a noun's "behavior," I do not intend to suggest animal or human behavior; but it will do no harm to think of it as analogous to the behavior of molecules, for instance. The laws of composition we shall touch on are not as precise as the laws of chemistry—they are closer instead to the rules of musical composition—but they are precise enough that the analogy between nouns and molecules is helpful. To ensure that it does not mislead, we shall be reminded at the conclusion of the article and in various places throughout that it is a poet and his technique that in fact determine how the nouns behave. Since we are interested in the nouns' *formulaic* behavior, let us begin by offering a precise definition of "noun-formulae," a definition suitable to the use of statistics: noun-verb and noun-epithet phrases that are exactly repeated, repeated with slight alterations (such as inflexion, separation, inversion, change in position, and extension), or partly repeated (the phrase contains a patronymic, or a generic word—an epithet or verb used in identical metrical circumstances with more than one different noun). Repetitions with alterations, and partial repetitions—inexact repetitions, that is—are counted as different formulae from those they inexactly repeat, so that it is possible

[1] In speaking of Homer as one author, as I shall do, I do not mean to imply anything more than that from the point of view of formulaic composition, I can detect no difference between the *Iliad* and the *Odyssey*. To make this clear, equations will be given for the individual poems as well as for Homer generally.

(and indeed common) for a formula to occur only once.[2]

The first correspondence between Homer and the *Roland* is simple enough: the nouns in both are consistently formulaic, and the nouns in one have about the same **formularity** (the same percentage of formulaic occurrences) as the nouns in the other: 74.8% in Homer, 70.5% in the *Roland*. When I say "consistently formulaic" I mean that though the formularities of some nouns in each source can vary considerably, Homer's mostly tend to cluster around 74.8%, the *Roland*'s mostly around 70.5%. This consistency is most clearly revealed when we construct linear equations relating **formulaic occurrences** to **total occurrences** (they are given below, in Section III). With some exceptions, the bulk of Homer's nouns that occur often enough for useful statistical comparisons display a value for **formulaic occurrences** that is very close to the value we expect from the equation (the expected or "predicted" value); and exactly the same is true of the *Roland*'s nouns.

When I say "about the same **formularity**" I mean that, despite the difference between 74.8% and 70.5%, the parameters—the slope and the y-intercept—of the Homeric equation are nearly identical to the parameters of the *Roland* equation. Hence we can feed figures for **total occurrences** from the *Roland* into the Homeric equation (or vice versa) and come up with figures for **formulaic occurrences** in the *Roland* (or Homer) that are very close to the truth. In other words, we can regard the parameters of the Homeric equation as a prediction, remarkably accurate, of the parameters of the equation for the *Roland*. To this extent the poets must share the same compositional technique: in his handling of nouns, each uses a formula about as often as the other, roughly three-quarters of the time; and each appears to aim consistently at this figure.

The second correspondence arises when we construct equations relating the **total occurrences** of a noun to the number of **different formulae** it displays. A formula is different from another if it does not repeat it exactly, which is why all repetitions with slight alterations, and partial repetitions, are counted as occurring only once (unless they are

[2] A full discussion of the criteria for a statistically appropriate definition of a formula is given in Sale 1989:347-51.

themselves repeated exactly).[3] We might have expected that the more often a noun occurred, the more often each of its different formulae would be used. Instead, we find that on the whole this is not true; rather, the more often a noun occurs, the more **different formulae** it generates, while the number of its **occurrences per formula** does not grow much.[4] We can construct a linear equation relating the number of **different formulae** to **total occurrences**, from which, if we know a noun's **total occurrences**, we can make a good calculation of its **different formulae**.

We can make an even better calculation by introducing two new concepts. First, **localization**, the percentage of times a noun occurs at the point at which it occurs most often. Some nouns, especially in Homer, tend almost always to be found in just one position in the verse (called the localization-point), while others wander about, and as they wander, create different formulae in various parts of the line. There is a limit to the number of formulae that can be generated from any one position, because the poets will not create different formulae that say exactly the same thing in exactly the same meter. Hence we expect, and find, that in both poets, nouns that wander possess more **different formulae** than nouns that do not, though their **total occurrences** may be exactly the same. Second, though **occurrences per formula** does not change much with **total occurrences**, it does change a little, and we find that by introducing it as another variable into the equation we can improve the calculation. This revised Homeric equation not only fits the Homeric data elegantly, it makes extremely accurate predictions for the parameters of the corresponding equation for the *Roland*. As a result we can feed figures for **total occurrences**, **localization**, and **occurrences per formula** for the *Roland* into the Homeric equation, and come up with remarkably accurate figures for

[3] The one exception to this is that *extensions* are not counted as different formulae, since if they were we would have to make some very bizzare statements: for instance, the formula λευκώλενος Ἥρη occurs only three times and is an infrequent formula (but θεὰ λευκώλενος Ἥρη occurs 19 times and is a frequent, a regular formula).

[4] To make such statements as "The more often a noun occurs, the more different formulae it generates," is somewhat sloppy; it suggests that we are following the course of a given noun through a number of poems. But this language is far handier than such locutions as, "When one noun occurs more often than another, the number of its different formulae will be proportionately greater than the number of different formulae displayed by the other noun."

different formulae for the *Roland*.[5] Here again we have demonstrated a shared compositional technique: roughly stated, the more often a poet uses a noun, the more different formulae he will employ for that noun; neither poet elects simply to use the same formulae more often. Later we shall find reason to connect this technique with oral composition in performance; but even without this inference it is interesting to uncover this shared rule of epic creation.

We note the third correspondence when we construct graphs, hyperbolic in shape, that pinpoint the difference in behavior between frequently-occurring formulae (what I call "regular formulae") and infrequent formulae. The x-axis of these graphs gives the number of times a formula occurs: once only, twice, three times, and so on. (Remember that a formula that is never exactly repeated, but repeats another formula inexactly, is said to occur only once, because it counts as a different formula from the one it inexactly repeats.) The y-axis gives the number of formulae that occur at each level on the x-axis: Homer, for instance, has 673 formulae that occur once only, 490 that occur just twice, and so on. There are many fewer formulae that occur twice than once, many fewer occurring three times than twice, many fewer four times than three times, five times than four times, six times than five times. The descent is steep and almost linear. But at 6 times (on the Homer graph) a change occurs: the plunge is arrested, and we find virtually as many formulae exactly repeated 22 times as 11 times (for example). This change of behavior, once we have analyzed it, enables us to identify a small range of numbers from which to choose a *minimum number* for a formula to be counted as frequently occurring, to be called a regular formula. It turns out that the choice of 6 is a reasonable one for the *Roland* as well as Homer. It is striking that hyperbolae occur in both authors—that is, that both authors distinguish between regular and infrequent formulae; it is more than striking, it is astonishing that both hyperbolae offer a similar range of choices, such that it is reasonable to pick the same number for our minimum in both authors.

In the course of studying the difference between regular and infrequent formulae, we observe that each Homeric noun displays only a

---

[5] I cannot sufficiently stress that what the Homeric equation is predicting is the parameters of the *Roland* equation, not the figures for its **different formulae**, which will of course be predicted accurately if the parameters are sufficiently close. It would be uncomfortable to speak of predicting **different formulae** with an equation that included **occurrences per formula** on its right-hand side, since we cannot know **occurrences per formula** until we know **different formulae**.

few different regular formulae, usually 1 or 2 (true for 87% of the 190 Homeric nouns studied), and almost always 4 or fewer (97%). Therefore, though a frequently-occurring noun may well have more different regular formulae than a noun occurring less often, it will obviously not have many more. This will hold down the number of different formulae a noun can display, and work against the general rule we stated above, that if a noun occurs more often, it will display more different formulae. On the other hand, a frequently-occurring noun does show proportionately more regular formulaic *occurrences*. As a result, there is a tendency (not remarkable, but genuine) for such a noun to show more occurrences *per* regular formula (this too working against the general rule). As **total occurrences** goes up, **occurrences per regular formula** tends to go up. This means that **occurrences per formula** will go up too—not much, to be sure, because there are many fewer regular formulae than infrequent formulae, and because **occurrences per *infrequent* formula** is very nearly constant with respect to **total occurrences**. But it goes up enough to explain why we need to make the slight modification suggested above to the equation relating **total occurrences** and **different formulae**.

These statements will grow clearer as we proceed; I have made them here in order to emphasize that in Homer most of a noun's formulaic behavior is absolutely regular, "statistically predictable." We have already seen that from its **total occurrences** we can determine its **formulaic occurrences** and its **different formulae**. The cap on the number of regular formulae makes it relatively easy to find out how many of these different formulae will be regular formulae and how many infrequent formulae; then, since **occurrences per infrequent formulae** is constant, we can equally easily discover how many of its **formulaic occurrences** will be regular and how many infrequent. We can do this because the overall formulaic technique is pervasive: it reaches into every corner of the poem. The same is true of the *Roland*, only here the influence of **occurrences per formula** and especially **localization** is much slighter, and we make our discoveries simply by determining **formulaic occurrences** and **different formulae** from **total occurrences**. Our ability to predict the formulaic behavior of each poet does not by any means suggest that either one was a mere mechanic. It means that each one followed a technique of composition, followed rules of procedure little different from the rules that musical composers follow. And if something is a rule of composition, it is usually obeyed throughout the piece. We find similarities in harmonic progression among most instances of the Classical sonata-allegro form, similarities that are no doubt susceptible of statistical analysis; the intellect of a Mozart utilizes the common technique even as it individualizes and deepens it. Thanks to the preservation of a large number of *chansons de geste*, we can observe the genius of Turoldus, the *Roland* poet, in his mastery of (and over) the technique he shares with his fellow *jongleurs*. The freedom enjoyed by Homer is even greater, partly because greater freedom is built into the metrics of the hexameter line. But there is nonetheless a body of strict rules that both poets obeyed.

Mathematics uncovers certain of these rules, and shows that they are the same for both poets. Both halves of this statement are equally important to us—that each poet *had* a mathematically discoverable technique for handling formulae, and that it was pretty much the *same* technique as regards the utility of formulae. We not only isolate algebraic equations, but we stress their similarity. We are glad to discover the hyperbola for Homer, since it confirms the distinction between regular and infrequent formulae; we are even more pleased to discover a hyperbola for the *Roland*, and to note how similar it is to Homer's. In what follows we shall first set out the equations and the hyperbolae, and then shall go on to try to explain these similarities, adducing the results of similar (not yet published) investigations of the *Cantar de Mio Cid* and Avdo Međedović's *Wedding of Meho, Son of Smail*, poems that behave in the same way as the *Iliad*, the *Odyssey*, and the *Roland*.

In giving this explanation I shall find no other way to account for such impressive resemblances except on the theory that the technique that generated them was created to facilitate the composition of *oral* poetry in *performance*. This is not just a conclusion *faute de mieux*, or a mere assertion that orality is the only thing they have in common that can possibly explain the mathematical similarities (though it is true that the rest of what they have in common does not explain them). Rather, we shall argue that the similarities are due to the poets' meeting certain of their needs by generating infrequent formulae. Such needs arise rarely for any given noun, but are of the sort that arise constantly; the formulae to meet them are generated either out of previously existing materials—generic words and alterable formulae—or by repeating exactly a phrase used earlier in the poem. This material is kept ready to hand, and turned into new formulae so predictably and so pervasively, because the demands of composition in oral performance are immediate and unrelenting: unusual needs arise at every turn, there is no leisure to investigate a variety of possibilities, and there is no rest until the piece is over.[6]

[6] We shall discuss below (Appendix 1) how unusual needs can "arise at every turn." This process of composition is exactly what Albert Lord describes as "adjustment of phrase and creation of phrases by analogy" (1960:37). Indeed, the distinction between

It is important to stress the limitations on the above argument. Mathematical comparisons show that in three important ways the techniques of the two poets were remarkably similar; only in attempting to explain the similarity do we have recourse to the theory of oral composition. Even here we must be careful. It is at least possible that any one of our poems (except Avdo's) was composed in writing; it is only the underlying *technique* that must be oral. We may wonder why a literate poet should have wished to reproduce so thoroughly the circumstances, and the results, of composition in performance; but we cannot disprove the possibility that he did. It is also at least barely possible that any one of our poems was preserved orally for a long period of time; the underlying technique bespeaks only an *original* composition in performance. The culture of the *Roland* had the means for writing the poem down at any time during the period in which it must have been composed; but we may have to suppose a period of oral preservation for the Homeric poems.[7] This is acceptable, from the point of view of our arguments, provided that the preservation was careful and the effects of composition-in-performance were not destroyed.

The best sustained demonstration known to me that the *Roland* is an orally composed poem is to be found in Joseph Duggan's *The Song of Roland* (1973; see also the excellent assessment of the value of this work in Foley 1988:79-80 and 96-97). Duggan approaches his task from several points of view, of which statistical analysis is only one part; indeed most of his book is given over to a qualitative discussion of how formulae work (both in the *Roland* and in certain other *chansons de geste*), of how they form the basis of the narrative technique, and of how the poet of the Oxford *Roland* employs that technique in composing such magnificent poetry. But

---

infrequent and regular formulae as based upon the hyperbola, as well as the equations (espec. 2A-C below) that are ultimately explained by this distinction, are little more than a mathematical formulation of the results of the process described by Lord on pp. 37-67.

[7] I must say I find Richard Janko's (1990) restatement and elaboration of Albert Lord's theory of oral dictated texts extremely plausible. I am also convinced by Janko's eighth-century date for the *Iliad*, which I have argued for on other grounds—see Sale 1987:38. But there are two outside possibilities that cannot be dismissed. First, as Gregory Nagy has pointed out (in conversation), the statistically determined linguistic differences among the various early epic poems might be due not to difference in time of composition, but to difference of place and tradition. Second, as Sarantis Symeonoglou has pointed out (also in conversation), the eighth century might have wished to write down the text of the *Iliad*, but have simply lacked a sufficient supply of material on which to write it. In that case, singers might have devised a technique of verbatim oral transmission quite different from the technique of composition in performance.

Duggan does not neglect the mathematical side: using rigorous criteria for what is to constitute a formula, he finds the poem 35.2% formulaic—that is, 35.2% of its hemistichs are themselves formulae. This makes it comparable in formularity to Old French poems known to be orally composed, and much more formulaic than a large body of medieval material known to have been composed in writing. Duggan's criteria for a formula are slightly different from mine: he does not include phrases that occur only once but share a key epithet or verb with a similar phrase of identical metrical shape and syntax. These he calls "syntactic formulas"; Lord calls them "formulaic expressions," and I call them "generic formulae" in order to direct attention to the shared key word; they are the phrases I classify as "partly repeated" above. But Duggan does count phrases that undergo modifications similar to the alterations set out and discussed by J.B. Hainsworth (1968:*passim*; my "slight alterations" above).[8] As we have already seen, I count both generic formulae and Hainsworth-alterations as formulae. On the other hand, I have occasionally refused to count certain repetitions confined entirely to lines close to one another, where there is a possibility of a refrain effect, of deliberate echoing—in other words, where the repetition may not be integral to the narrative technique.[9] Nor do I count such phrases as "Li empereres" filling

[8] See esp. Duggan 1973:131-33. The alterations noted by Duggan include: inflection, simple stylistic variation (such as "en cest pais" for "en ceste tere"), changes obviously brought about by metrical considerations ("cinquante carre" in the first hemistich becoming "plus de cinquante care" in the second), and changes in second-hemistich formulae due to assonantal requirements.

[9] See also Sale 1989:347, where some other sorts of repetition are also ruled out as sufficient criteria for formulae. There is some inevitable subjectivity here, since the decision that a passage is a deliberate echo is an aesthetic judgment; I have deliberately erred in all cases in favor of declaring that a passage *is* formulaic, but mistakes will surely occur. However, the number of instances where the problem arises is relatively small: the most I have observed for any one noun is three, and that many only when there are a great many total occurrences; most nouns present no problems at all. Since the total number of nouns, formulaic occurrences, and different formulae is so large, this source of error cannot significantly affect the statistics. Duggan uses small capitals when he prints formulae whose occurrences are confined to a given scene. This procedure allows us to recognize the possibility of refrains, while also identifying formulae that may have limited usefulness, or may have been coined for the sake of composing a particular scene. See Duggan 1973: 42.

Another seeming source of error is worth mentioning here: when a phrase is the first hemistich, phrases possessing neither an epithet nor a verb—phrases I call "minimal formulae." My criteria should by no means be seen as an implied criticism of Duggan's. Mine were created for Homeric analysis undertaken long before I had studied Duggan's work carefully; I continue to employ them because we are now looking for what Homer and the *Roland* have in common, and it behooves us to use the same criteria on both sets of data. In any case, what I call a formula and what Duggan calls a formula will on the whole coincide. I have given in Appendix 2 the phrases that are formulaic by my criteria for 22 of the nouns in the *Roland*.

The mathematical relationships I am exploring derive from my earlier work on Homer, and are different from those discussed by Duggan (see Sale 1984, 1987, and espec. 1989). To restate more formally what was set out earlier: the **total occurrences** of a given noun in a given grammatical case are grouped into a set, which is divided into subsets labeled **formulaic occurrences** and **non-formulaic occurrences**. The **formulaic occurrences** (algebraically, **fo**) of a noun divided by its **total occurrences** (**to**) is its **formularity**. We determine the relative formularity of the sets by constructing the linear equation **fo = ƒ(to)**, **formulaic occurrences** as a function of **total occurrences**. We then count the number of **different formulae** in each noun's set, and construct the linear equation **df = ƒ(to)**, **different formulae** as a function of **total occurrences**. We then calculate the **localization** (abbreviated **loc** in the equation) of each noun, and observe that **df = ƒ(loc)**, that **different formulae** is a function of **localization**. This fact justifies us in using **loc** to modify the equation **df = ƒ(to)**, producing **df = ƒ(to, loc)**. We then calculate **occurrences per formula (tof/df)** for each noun, and construct the linear equation **df = ƒ(to, fo/df)**. We go on to combine this with **df = ƒ(to, loc)** to produce the further modification **df = ƒ(to, loc, fo/df)**. Then we classify the different formulae for all our nouns, as to whether they occur once, twice, three times, and so on, and count how many occur once, how many twice, and so on. This gives us the distribution of number of occurrences for each level of frequency of occurrence, enabling us to construct the hyperbola and determine a minimum number for regular formulae. We are then able to divide each noun's different formulae into regular formulae and infrequent formulae, and its total formulaic occurrences into regular formulaic

---

repeated just once, even though it is almost certainly not a deliberate refrain, it might be repeated by accident. Here there is no real problem provided we are consistent. All such cases are rigidly counted as formulae for all nouns in all poems, so that the validity of the *comparisons* is not affected. The worst that can happen is our forming the opinion that the poets are slightly more formulaic than they really are.

occurrences and infrequent formulaic occurrences. This distinction is shown to underlie the **df = *f*(to, loc, fo/df)** equation and to account for the shape of the hyperbola; it is our ultimate basis for arguing that the technique of the two poems was developed for the sake of composition in performance.

II. Data and Definitions

In what follows we shall be comparing three groups of data:

1. The 190 nouns in Homer that occur at least 13 times and have at least one regular formula. A noun in the *Odyssey* is counted as a different noun from the same noun in the *Iliad*. We have three reasons for counting this way: some nouns occur in only one poem, and should not therefore suffer statistically; we need to keep the two poems independent in order to observe significant statistical differences between them, if any such arise (so far I have not encountered any); and we must be alert to the possibility that the length of a given poem might influence the statistics.

2. Twenty-two nouns in the *Roland* that occur at least 13 times and possess a regular formula: the 11 personal names in the nominative that occur this often (all 11 happen to possess a regular formula), plus 11 common nouns.[10] There are a few more common nouns, not many, that meet these criteria and might have been included; but I felt that there was a statistical advantage to having the same number of both types, and constructed a similarly divided set for Homer (#3 below). With a minimum of 12 I could have included "Blancandrins" (see note 41 below), but much experimentation with nouns in Homer had already convinced me that bringing the minimum below 13 brought about misleading improvements in the statistics.

We cannot, to be sure, claim that 6 is the only possible minimum for regular formulae in the *Roland* (we might have chosen 4, 6, or 8: see below), but let us at least note that all the personal names (not counting "Deus") occurring in the nominative at least 13 times (or 12 times, for that matter) possess a formula exactly repeated at least this often; no character is

---

[10] It is important to keep in mind that a phrase, in order to be counted as a regular formula, must be exactly repeated (disregarding certain irrelevant spelling variations and *verb* inflections): each part must fall in the same place in the line of verse, and the noun must be in the same case and number. "Blanche barbe," e.g., is not a regular formula, though the two words are juxtaposed 10 times. Some nouns—"sire," e.g.—are best treated as epithets, though they can be used independently.

being left out. Should later investigation reveal idiosyncratic behavior on the part of some common nouns omitted because they have no formula exactly repeated 6 times, that will bring complications, not falsification.

3. A smaller selection of 70 Homeric nouns in the nominative used 15 times or more and possessing at least one regular formula: specifically, the 35 personal names that in fact meet these criteria, and 35 common nouns chosen from among those meeting these criteria.[11] This selection was constructed for two reasons:

First, I wanted something closer in size to the *Roland* set than the 190 nouns under #1 above; I wanted it exactly divided into proper and common nouns, like the *Roland* set; I wanted a minimum number of occurrences per noun that would keep the set relatively small while ensuring that some of its nouns occurred only a little more often than the *Roland* minimum; and I wanted to avoid as much sample bias as I could—that is, I wanted to minimize my own choosing of the individual nouns that were to belong to the set. The minimum of 15 occurrences per noun determined 35 personal names automatically, so that if I let the set have 70 nouns, divided half common and half proper (as the *Roland* is divided), then half the set, at least, could be unbiased. A minimum of 13 would have produced too large a set, while a set the same size as the *Roland* set would either have been subject to intolerable bias (I would have had to choose every member), or have entailed a much larger minimum number than the 13 for the *Roland*.

Second, I wanted a selection whose formulae could be broken down into those that fall in a major colon (see the definition below) and those that

---

[11] The choice of common nouns was on the whole random. It resulted in a rather larger than normal number of nouns with exceptional formularities; this fact did not affect the statistics importantly. I did deliberately avoid one noun, θυμός (appearing in three forms, θυμός in the *Iliad* and the *Odyssey*, and θυμόν in the *Iliad)*, whose formulae include one that is metrically bizarre by Homeric standards: θυμὸς(ν) ἐνὶ στήθεσσι. Including all three appearances would not have compromised either the statistics or the argument, but it would have made the statements much more cumbersome: I would have been constantly interrupting an already difficult presentation to remind readers that this isolated formula, which was producing some minor ripples on the graphs, really is isolated; and I would have had to say in nearly half-a-dozen places, "except, of course, for θυμὸς(ν) ἐνὶ στήθεσσι." That this formula does not really affect the fundamental statistics is clear from its presence (three times, of course) among the 190 nouns, where its distorting effect is virtually unnoticeable. (Granted, one of the three cases, θυμός in the *Iliad*, is a distinct outlier for two of our equations: it is too formulaic and has too many different formulae. But this means that some of what I counted as infrequent formulae are probably chance repetitions, an experimental error bound to occur sometimes with nouns occurring over 100 times.) And subtracting it from the group produces very little change.

do not, in order to explain the hyperbola and to make comparisons with *Roland* formulae that fall, or fail to fall, in the first hemistich. To do such a breakdown carefully for 190 nouns is a monumental enterprise, and before engaging in it I wanted to discover whether comparisons along these lines between smaller groups of sets would prove fruitful.

The following definitions, evolved from Homer, have also been applied to the *Roland*.

Caesura = a break in the poetic line after a word-ending. The caesurae that most interested Milman Parry fell after verse-positions 5, 5.5, 7 and 8 (see below) and are called, respectively, the penthemimeral (masculine, B1), trochaic (feminine, B2), and hepthemimeral (C1) caesurae and the bucolic diaeresis (C2). The caesura in the French decasyllabic line always falls after verse-position 4 or 4.5.

Colon = a segment of the hexameter line falling between two caesurae or between a caesura and the beginning or end of the verse.

Epithet = an adjective, adjective-phrase, noun, or noun-phrase accompanying a noun in a formula.

Formula = a noun-epithet and noun-verb phrase, either

A. exactly repeated (same words, same grammatical case, same place in the line of verse), or
B. repeated with slight variations (Hainsworth-alterations), or
C. partly repeated by including a generic epithet or verb so as to constitute a generic formula, or
D. partly repeated by including a patronymic.

Formularity = formulaic occurrences ÷ total occurrences.

Generic epithet or noun = an epithet or verb used in identical metrical circumstances with at least two nouns of the same metrical shape.

Hainsworth-alteration = a formula that differs from another merely by occupying a different position in the verse, or being extended by an added word, or by being inflected, or by having its parts separated or inverted.

Hemistich = a segment of the hexameter line running from the beginning though verse-positions 5 or 5.5 (occasionally 6) or from 5.5 or 6 (occasionally 7) to the end; a segment of the French decasyllable occupying the space before, or the space after, the caesura.

Infrequent formula = a formula exactly repeated fewer than 6 times, or (if it occurs only once) containing a generic epithet or verb, or a patronymic, or consisting of a Hainsworth-alteration.

Localization-point = the place in the verse in which a word occurs most often.

Localization = the percentage of times a word occurs at the localization-point.[12]

Major cola = the cola in the hexameter line that run from the beginning to verse-positions 5 and 5.5, from 5, 5.5, 7 and 8 to the end, and from 2 or 3 through 8.[13]

Minimal formulae = single words, and noun-preposition and noun-adverb phrases, that fall repeatedly in the same place in the line.

Minor cola = all cola except major cola.

Regular formula = a formula exactly repeated 6 times or more in any one poem.

Regularity = regular formulaic occurrences as a percentage of formulaic occurrences.

Verse-position = 1) a segment of the hexameter line occupying one long syllable or two short syllables and numbered from the beginning of the line. Thus position 1 is the opening long syllable, position 1.5 (or 1 1/2) the ensuing short syllable if there is one, position 2 the second long syllable, and so on. 2) a segment of the decasyllabic line occupying one syllable and numbered from the beginning of the line. Ten syllables is normal; but after 4 and 10 we may have 4.5 and 10.5.

I have put two Appendices at the end of the article. The first is a discussion of how infrequent formulae come into being, a discussion that seemed too elaborate for the text itself. The second gives all the data for the *Roland*, along with a list of its regular formulae. Some of the data for Homer are published in Sale 1989:396-405; the rest are fairly easy to compile with the help of the concordances, or the Ibycus computer, or the Pandora program for the Macintosh, using the same format that I used for the *Roland* in Appendix 2. But I would be happy to respond to individual requests.

---

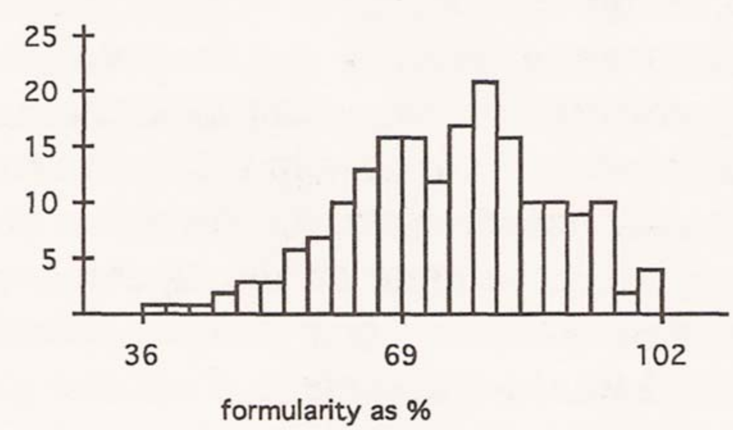
[12] Though it is natural to state this as a percentage, it is desirable in constructing equations that all the variables have comparable sizes, and this desideratum has entailed stating localization as a number from 1 to 10 followed by a decimal in Equations 3A, B, and C below.

[13] A full recent account of the caesura and the colon in relation to the formula can be found in Foley 1990:73-84. It should be noted that twice as many major cola fall after the caesurae as before them; it is also true that many more formulae fall in these second-half cola than in the first-half variety. This is in keeping with the principle that Foley calls "right-justification," the tendency for greater fixity at the end of the line (see Foley 1990:56-57, and below).

III. The Formularity Curve (Total Formulaic Occurrences/Total Occurrences)

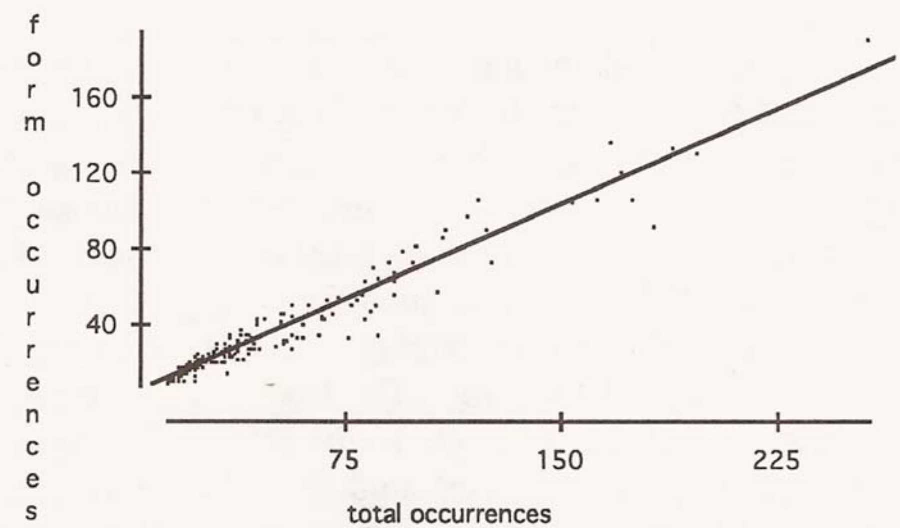
We begin with a fundamental fact: that Homer and the *Roland* both maintain consistent formularities for most of the nouns they use. There are two mathematical ways to state this consistency. One can construct a histogram for each poet, a picture of the distribution of the **formularities**, of the quotients **formulaic occurrences÷total occurrences**, of each of his nouns. These **formularities** cluster around the *average* formularity (74.8% in Homer, 70.5% in the *Roland*) in roughly bell-shaped curves. Figure 1 is the histogram for Homer's 190 nouns. The base of each rectangle spans 3 percentage points, so that the rectangle to the left of 69% means, "All the nouns with formularities between 66 and 69%." The y-axis tells how many: in this case, 16 nouns have such formularities. Note that the longest rectangle falls between 78 and 81%, higher than the mean of 74.8%; we do not have a precisely normal distribution, but one with a slight skew to the left.

Figure 1



Alternatively, one can state the algebraic relationship between the **total occurrences** of all the nouns and their **formulaic occurrences**. We derive this relationship by a process of *linear regression*. That is, we first construct a graph with **total occurrences** on the x-axis and **formulaic occurrences** on the y-axis, and put on it a point for each noun corresponding to that noun's **total occurrences** and f**ormulaic occurrences.** We then determine (by the method of least squares, which any statistical computer program will employ) what straight line comes closest to the points, gives the best fit. Figure 2 is the graph for 190 nouns in Homer:

Figure 2



The point furthest to the right, with 256 **total occurrences** and 191 **formulaic occurrences**, is Odysseus. The reader can see that the two methods, the histogram and the graph, correspond conceptually. That is, the line on the graph, though not precisely identical with it, is roughly equivalent to the mean formularity on the histogram; the distance between the line on the graph and the various points corresponds roughly to the distance between the mean formularity and the points in each rectangle in the histogram. The correspondence is not exact because the line on the graph reflects the fact that the variations in **total occurrences** of the nouns can affect their **formulaic occurrences**; the histogram omits this fact.[14]

We shall therefore concentrate on the algebraic relationship, which we shall need in any case when we come to study the relationship between **different formulae** and **total occurrences**. First, the 190 nouns in Homer (**fo = formulaic occurrences, to = total occurrences**):

1A. fo = .676 to + 2.1; R = .97, s = 6.9.[15]

---

[14] If we know the difference (call it **Δf**) between the formularity of a given noun and the mean formularity of all the nouns, .748 (stated as a decimal, not a percent), and we wish to know $\mathbf{s_1}$, the distance on the y-axis between the line and the point on the graph occupied by the noun, we use the formula $\mathbf{s_1} = \mathbf{to}(\mathbf{\Delta f} + .072) - 2.1$. That is, $\mathbf{s_1}$ includes the effect of **to**, **Δf** does not.

[15] The letter **R** stands for the correlation coefficient, one measure of the probable accuracy of predictions made from the equation upon fresh data. Since **R** = 1 means that the correlation is perfect, and **R** = 0 that there is no correlation, an **R** value of .97 marks a

The equation indicates that on average, the **formulaic occurrences** (**fo**) of a noun in Homer will be a little more (2.1 occurrences more) than about two-thirds (.676) of its **total occurrences** (**to**). Next, the corresponding equation for the 70 nouns in Homer, adduced mainly in order to affirm the fact that these 70 nouns are indeed a representative sample:

1B. fo = .679 to + 2.6; R = .98, s = 7.7.[16]

The reader will note that the equation is virtually identical to Equation 1A. Finally, the equation for the 22 nouns in the *Roland*:

1C. fo = .689 to + 0.3; R = .98, s = 4.7.

This equation is very close to the other two. If a noun occurs 100 times, Equation 1C predicts that it will have 69.2 formulaic occurrences; if a Homeric noun occurs 100 times, Equation 1A predicts that it will have 69.7 formulaic occurrences. And the biggest difference in predicted formulaic occurrences between the two equations is less than two.

These equations, then, give us the ratio between formulaic occurrences and total occurrences: the more often a noun occurs, the more formulaic occurrences it will have; and the difference between one noun and another in this respect is proportional. The correlation coefficients, **R**, are very high indeed at .97, .98, .98. The root-mean-square residuals, **s**, are low or reasonably low at 6.9, 7.7, 4.7. If we take the expected values (the "predicted" values) of **formulaic occurrences** from Equation 1A, Homer's 190 nouns, and compare the actual values, we find a low median error of 2.9, a fairly low average error of 4.6, but a maximum error of 33. This

---

very high correlation. The letter **s** stands for the root-mean square residual, an indicator of how far away such predictions will be from the observed or actual values. The median error that the equation in fact incurs for the Homeric data is 2.9; the median number of total formulaic occurrences displayed by our nouns is 25, so that the median error is about 11%. Given all the things that can affect a noun's formularity—its metrical properties, its meaning, the variety of contexts in which it may be used—a median error of 2.9 is satisfactory. (For the relationships among meter, meaning, and formularity, see Sale 1989:357-61.)

[16] The equations for the *Iliad* and *Odyssey*:

Iliad:	**tfo = .650 to + 2.9, r = .96, s = 7.7**
Odyssey:	**tfo = .726 to + .65, r = .98, s = 5.3**

means that most of the time the equation gives a highly satisfactory picture of the relation between **total occurrences** and **formulaic occurrences** in Homer, but that some nouns are really quite deviant. The *Roland* shows a comparable figure for the median error (2.7) but a lower average (3.4) and maximum (12).

As we saw above, the average formularity for Homer is 74.8%, for the *Roland* 70.5%, which means that if we apply the Homeric equation, Equation 1A, to the data in the *Roland*—that is, if we calculate **formulaic occurrences** for the 22 nouns in the *Roland* by feeding their **total occurrences** into the *Homeric* equation—we expect the calculation to be a little high; and so it is, by an average of 1.2. Despite this fact, it is still very close: the median error in its predictions is 2.8, the average 3.6, the maximum 11.[17] These results are almost the same as what we obtain from Equation 1C, the *Roland* equation, itself: a median error in *its* predictions of 2.7 (an improvement of just .1), average 3.4 (an improvement of .2), maximum 12 (not as good, by 1). Above all, Equation 1A has given a highly accurate prediction of the *parameters* of 1C: the slopes (.676 and .689) are very close, differing by only .01, while the y-intercepts (2.1, .3) are off by only 1.8 (that is, a difference of 2 formulaic occurrences). This means that the nouns in both sources are displaying the same *consistency*, are clustering near the mean formularity, or deviating from it, about as frequently and to about the same extent. Each shows a roughly normal distribution around the mean; Homer's standard deviation is 12.9, the *Roland*'s 11.4. The difference in average formularity of 4.3% between Homer and the *Roland* does not obscure the fact that both poets are using the same technique with regard to the formularity of their nouns.

IV. The Number of Different Formulae

The precision of the formularity relationship—the proportionality with which **formulaic occurrences** rises and falls with **total occurrences** both in Homer and in *Roland*—leads to a further conclusion. If a particular noun has more **total occurrences**—and therefore more **formulaic occurrences**—than another, it must *either* have more **different formulae**

[17] The calculated value and the amount off: Charles, 92(5), Roland, 83(11), Guenes, 39(3), Oliver, 29(2), Naimes, 17(4), Marsilie, 33(2), L'arcevesque, 24(7), Baligant, 13(3), Franceis, 41(7), Franc(s), 29(4), Paiens, 40(2), cheval 26(1), escut 14(1), hanste 12(4), osberc 18(3), reis 53(1), mot 19(0), cors 47(10), rei 31(3), cumpainz 12(1), bataille 22(1), oilz 15(2).

than the other, *or* its formulae must on average occur more often—it must display more **occurrences per formula**—or *both*. Earlier we stated the general rule, that the number of **different formulae** rises and falls with **total occurrences**, and stressed that this result is not trivial, that we can easily imagine a technique in which it was the other way around. Indeed it probably would have been if all we had were **regular formulae**, since **occurrences per regular formulae** does go up and down with **total occurrences**, and over the years I have noted a number of statements by scholars that seemed to imply a belief, perhaps half-conscious, that as the number of **total occurrences** rose, **occurrences per formula** rose with it. The fact is that **occurrences per formula** is close to being constant.[18]

If it were absolutely constant, we could deduce the relationship between **different formulae** and **total occurrences** from the formularity equation. **Occurrences per formula**, remember, is **formulaic occurrences** divided by **different formulae**, algebraically **fo/df**. If this were constant, we could write **fo/df = K**; multiplying through by **df**, we get **fo = dfK.** Substituting **dfK** for **fo** in Equation 1A we get **dfK = .676to + 2.1**; dividing through by **K** we could then write:

P. df = .676to÷K + 2.1÷K.

(I call this equation "P" to indicate that it is a derived equation, not directly based on linear regression as 1A is.) Equation P states that if **occurrences per formula** is constant, a change in **total occurrences** is accompanied by a change in **different formulae** precisely proportionate to the change in **formulaic occurrences** stated by Equation 1A.

Since **occurrences per formula** is not quite constant, we shall proceed a little differently. Because **fo = df(fo/df)**, we write Equation 1A as:

---

[18] Not quite, because as we just saw, **occurrences per regular formula** changes when **total occurrences** changes, and so therefore does **occurrences per formula**, at least slightly. It would not have to, if **occurrences per infrequent formula** (occurrences per infrequently employed formula) went down when **occurrences per regular formula** went up, but **occurrences per infrequent formula** does not; there is literally zero correlation between these two variables, and **occurrences per infrequent formula** is essentially constant. On the other hand, the correlation between **occurrences per regular formula** and **occurrences per formula** is quite good (.65 correlation coefficient); when one rises with total occurrences, the other does. When they do, **different formulae** is somewhat lower than it would have been had **occurrences per formula** been absolutely stationary, and we shall work this fact into Equation 4A.

Q. df = .676(to÷[fo/df]) + 2.1÷fo/df.

Since **df** can be canceled out—**df** could be anything without affecting **to** or **fo**—we cannot use Equation Q to determine the relationship between **total occurrences** and **different formulae**. But since **occurrences per formula** is *nearly* constant, we could guess the relationship between **total occurrences** and **different formulae** by entering the *average* value of **occurrences per formula** (3.898) into Equation Q. If we do this, we get

R. df = .173to + .54.

This is not far off the equation we get when we simply apply linear regression, the method we used to construct Equation 1A, to the data (see Equation 2A below); Equation R produces predicted values for **different formulae** that are virtually as close to the actual values as those predicted by Equation 2A. Equation Q therefore tells us that since **occurrences per formula** is *nearly* constant, a change in **total occurrences** is accompanied by a change in **different formulae** *roughly* proportionate to the change in **formulaic occurrences** stated by Equation 1A.

It also says that the slight changes that do take place in **occurrences per formula** could affect the relationship between **different formulae** and **total occurrences** *inversely*. If, say, **occurrences per formula** is higher when **total occurrences** is higher, **different formulae** will be not as high as it might otherwise have been. We were ready for this. We began Section IV by noticing that as **formulaic occurrences** changes, either **different formulae** or **occurrences per formula** or both must change, and change inversely: the greater the change in one, the slighter the change (or the greater the inverse change) in the other. This is logically necessary: it follows from the meaning of the concept **formulaic occurrences**. To this logical observation we add the empirical observation that when **total occurrences** changes, there is a corresponding change in **formulaic occurrences** (Equation 1A). It follows that as **total occurrences** changes (and **formulaic occurrences** along with it), either **different formulae** or **occurrences per formula** or both must change, and change inversely to each other: the greater the change in one, the slighter the change (or the greater the inverse change) in the other.[19]

[19] Note that the fact that we can cancel out **df** from Equation P does not make the equation a tautology, as it would be if the independent variable were **fo÷fo/df**. It is a restatement of 1A and says as much as 1A says. *If* we know a noun's **occurrences per formula** and **total occurrences**, we can determine first its **formulaic occurrences** and then

Let us set Equation Q aside for a moment, and turn to the equation relating **different formulae** (algrebraically **df**) and **total occurrences** as determined by linear regression from the data for the 190 Homeric nouns:

2A. df = .150to + 2.48, R = .83, s = 4.1.

That is, if the number of total occurrences of a noun is higher than another's, it will probably display more different formulae in the ratio indicated by the equation. Similarly for the 70 Homeric nouns:

2B. df = .120to + 3.78, R = .76, s = 5.2

And for the *Roland*:

2C. df = .236to - 0.06, R = .97, s = 2.0

The correlation coefficient for Equation 2C is very high; but the fit of the Homeric equations to the data, though all right, is not impressive, nor do the Homeric equations resemble the *Roland* equation as closely as we should like. Moreover, if we feed data from the *Roland* into 2A, we are off by an average of 2.3 and a maximum of 12.6, which is quite high.[20]

The *Roland* equation is so much more successful, indeed, that from it alone we might (begging the question for a moment!) suspect that something is missing from the Homeric equation, that it needs to be modified. We saw above why **different formulae** moves in inverse proportion to **localization**: if a noun is more highly localized, it will show a tendency to display fewer different formulae, because whenever it occurs at the localization point, it will very often use a formula it has already used

---

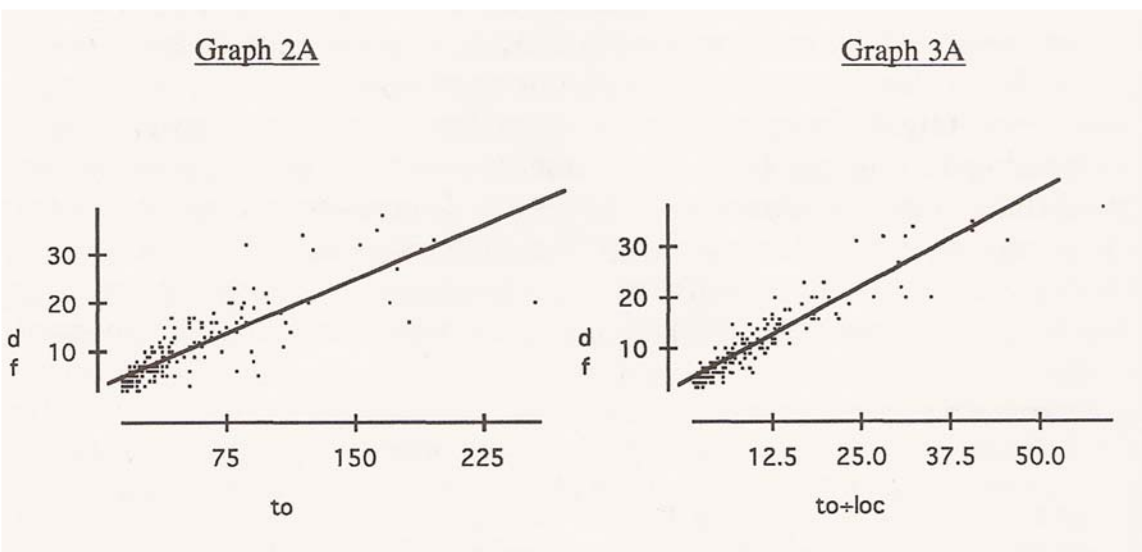
its **different formulae**, if and only if it has the formularity that Equation 1A says it should—which is to say, if and only if the parameters of Equation P are correct. Or we could know **total occurrences** and **different formulae**; we still do not know **occurrences per formula** unless we know **formulaic occurrences**, and we cannot figure that out from **total occurrences**—unless the parameters of Equation P are correct. If the parameters of P are correct, then a change in **total occurrences** will necessarily be accompanied by a change in either **occurrences per formula** or **different formulae** (or both), but we cannot know which one *a priori*.

[20] High, because the range of values for **different formulae** is much smaller than for **formulaic occurrences** in Equation 1A above; **different formulae** in Homer goes from 2 to 38, while **formulaic occurrences** goes from 8 to 191.

before.[21] The reason for this is the principle of economy: Homer has almost no formulae with the same referent and the same meter, unless the sense is genuinely different. There is obviously a limited number of formulae that can put the name at the localization point and still differ from one another. Hence a noun that is highly localized eventually faces the choice of repeating a formula already used, or of violating economy, and it rarely prefers the latter. The equation expressing the relationship between **different formulae** and **localization** has therefore a negative correlation coefficient, expressing the inverse proportion. The value is -0.50, not high; but for such a large sample size (190) the correlation is certain. Hence it is logical to combine **localization** (algebraically **loc**) with **total occurrences** in a new variable **to÷loc** (we put **loc** into the denominator since it moves inversely with **different formulae**) and perform the linear regression for **df** and **to÷loc**:

3A. df = .752(to÷loc) + 2.9, R = .93, s = 2.7

In order to indicate the extent of the improvement of 3A over 2A, I give graphs for the two:



[21] ʼΟδυσσεύς (with double sigma) in the *Odyssey*, for instance, occurs more often (256 times) than any other noun, but his figure of 20 different formulae is equalled or bettered by no fewer than 20 other nouns, some of which are found far less frequently. Ζεύς in the *Odyssey* occurs just 87 times, a third as often, yet has 32 different formulae, more than half again as many. Ζεύς, as a monosyllable, has low localization and wanders all over the verse; ʼΟδυσσεύς is highly localized.

The equation for the 70 nouns, affirming that this sample is typical:

3B. df = .702(to÷loc) + 3.3, R = .95, s = 2.6

These improved equations are matched by a comparable equation for the *Roland*:

3C. df = .950(to÷loc) + 1.8, R = .95, s = 2.5

The parameters of 3A and 3C are a little different, and the maximum error calculated by 3A for the *Roland*'s **different formulae** is 8, higher than we could wish. Still, the mean error is only 2, and the median still lower at 1.4. These equations, taken together with 2C for the *Roland*, are perfectly satisfactory evidence that the *Roland* and Homer are alike in this aspect of their techniques, that in both poets the more often a noun occurs, the more different formulae it generates.

I am, however, convinced from our examination of Equations P and Q that it is appropriate to add the complex variable **formulaic occurrences** divided by **occurrences per formula**, **to÷fo/df**, to 3A. It not only brings in the fact that earlier, by constructing this variable, we came close to deducing the **different formulae/total occurrences** relationship from 1A, but it also recognizes what we saw at that point: that since slight changes in **occurrences per formula** must have an effect upon the relationship between **different formulae** and **formulaic occurrences**, they also have an effect upon the relationship between **different formulae** and **total occurrences**. The result:

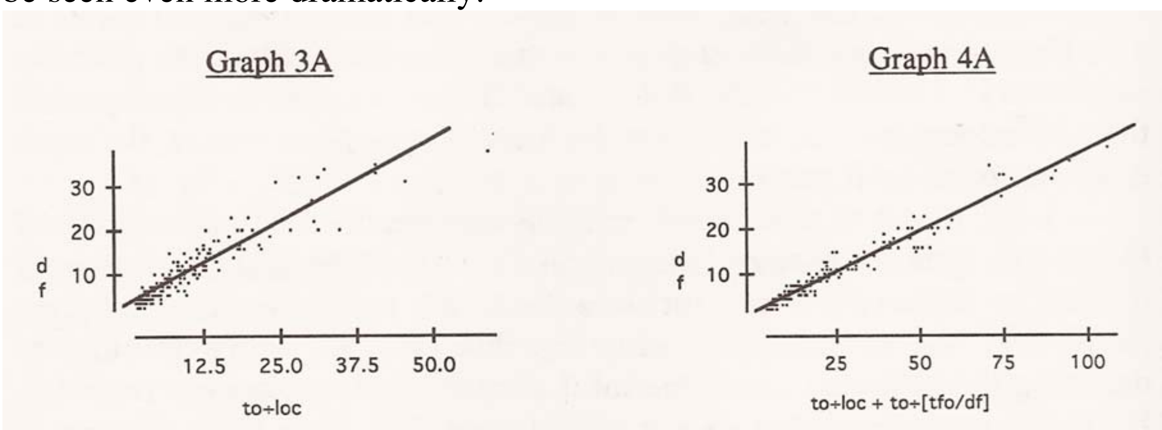
4A. df = .363 (to÷loc + to÷[fo/df]) + 1.4, R = .98, s = 1.6[22]

Localization and **occurrences per formula** are both in the denominators because both are in inverse proportion to **different formulae** (see above on **occurrences per formula**, and on **localization**).

---

[22] This equation should replace the one I gave in footnote 45 of Sale 1989:394; and we should modify somewhat the final statement in that footnote, which reads, "as TO rises and falls, the number of different formulae is affected precisely, but the occurrences per formula not at all." Occurrences per formula *is* affected by rises and falls in the total occurrences, though only very slightly. And the trouble with the equation I gave is that, though it has a very high correlation coefficient, it depends in part upon a correlation between infrequent formulaic occurrences and the number of different formulae, and this is a separate phenomenon with a separate explanation.

Equation 4A obviously gives us an extremely good fit. We must stress, though, that a full analysis and justification of it is still needed, that we do not have an entirely satisfactory mathematical statement of the relationship between Equation 1A and 4A, and of the precise role of the effect in 4A of the slight variations in **occurrences per formula** on **different formulae**. We have an equation which works *empirically*, and whose empirical logic we understand intuitively. The correlation coefficient, **R**, is higher than it is in 3A, and **s**, the mean residual, lower; if we apply Equation 4A to Homer's nouns, the average error is only 1.1. And when we compare the graphs of Equation 3A and 4A, the improvement can be seen even more dramatically:



I give the corresponding formula for the 70 nouns, again in order to validate the typical nature of the sample:

4B. df = .354 (to/loc + to÷[fo/df]) + 1.6, R = .98, s = 1.4[23]

Equation 4A gives an excellent prediction of the parameters of the corresponding *Roland* equation:

4C. df = .391 (to/loc + to÷[fo/df]) + .9, R = .98, s = 1.7

The parameters of, and the **R** and **s** values for, Equations 4A and 4C are much closer to each other than the parameters and values of 2A and 2C. And if now we feed *Roland* data into Homeric equation 4A, the mean error is 1.2 and the maximum only 4.5, a striking improvement over the results of applying 2A. The two variables **localization** and **occurrences per**

---

23 The equations for the *Iliad* and *Odyssey*:
Iliad: **df = .350 (to/loc + to÷tfo/df) + 1.6, r = .98, s = 1.7**
Odyssey: **df = .381 (to/loc + to÷tfo/df) + 1.1, r = .97, s = 1.4**

formula are obviously very meaningful for Homer. They are not so much so for the *Roland*, probably because the correlation between the two variables **different formulae** and **total occurrences** is already very high (coefficient .97). Nevertheless, 4A and 4C are very close to one another; and 4A, whatever its theoretical deficiencies, is an equation derived from Homer that gives an exceedingly accurate picture of the relationship between **different formulae** and **total occurrences** in the *Chanson de Roland*.

The compositional techniques are therefore in some respects the same. In both poets, nouns that occur more often have proportionately more different formulae than those occurring less often. The number of different formulae per noun is about the same in both poets: each poet's nouns display a minimum of 2; Charles has 35, Zeus in the *Iliad* also has 35, and νῆας in the *Iliad* has 38. Hence we can say: not just proportionately more different formulae, but significantly more. In both poets, there is a cap on the number of regular formulae, so that these significant and predictable differences in the number of different formulae are mostly differences in the number of infrequent formulae.

We cannot sufficiently stress the role played by the distinction between regular and infrequent formulae in creating these equations, and in marking the great similarities and slight differences between the techniques. In both poets, when a noun's **total occurrences** is high, its regular formulae show more **occurrences per regular formula**, while its infrequent formulae show more **different infrequent formulae**. If one or both poets had used a technique whereby **different formulae** was the same for nouns with low **total occurrences** and nouns with high, but **occurrences per formula** was very different, then the nouns in that poet's works could not have displayed very many infrequent formulae. Most formulae would have been regular formulae; but as it is Homer has five times as many infrequent formulae as regular formulae, the *Roland* seven! In both poets, the ratio **occurrences per formula** is nearly constant with **total occurrences**; in Homer this is clearly because **occurrences per *infrequent* formula** is constant (the cause is less certain in the *Roland*.) **Localization** in both poets moves inversely with **different formulae**, because when a noun has low **localization** and wanders about in the line, it is infrequent formulae that are generated in the unusual positions.

But although the distinction between infrequent formulae and regular formulae is highly significant, we know it so far only as a quantitative distinction. The step we are to take next will eventually reveal differences in quality.

V. The Formulae-Occurrences Curve in Homer

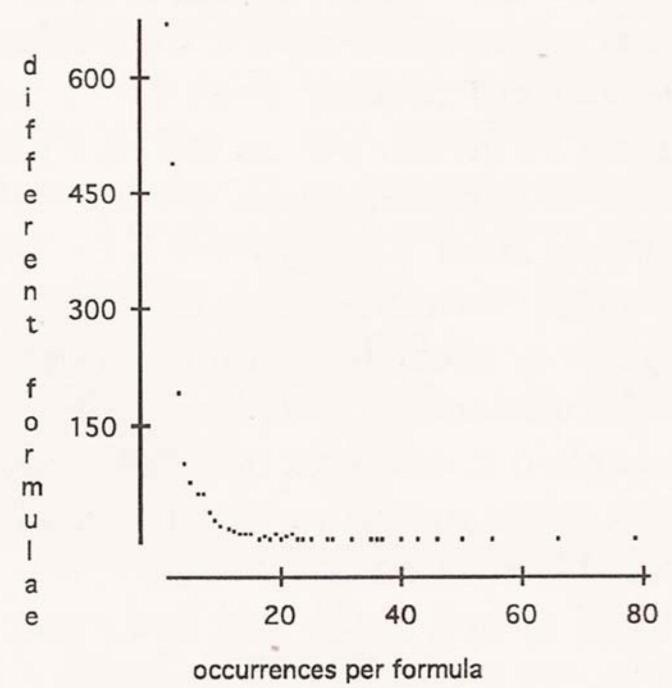
The distinction between regular and infrequent formulae is vital in Homeric studies for two reasons: first, for the reason we have just seen, that the equations that predict variations in the number of **different formulae** are largely based on variations in the number of **different infrequent formulae**; and second, because Parry's assertions about formulaic systems in Homer hold good for regular formulae and break down for infrequent formulae. I have discussed the second reason in earlier work, in which I also develop criteria for applying the term "regular formula" to the proper nouns in the nominative case in Homer, and defend the choice of 6 as a minimum number of occurrences while calling attention to the fact that 8 and 10 are also defensible minima (1989:362-95). I made the choice of 6 not because it was any more logical than the other two, but because I wanted to make it as low as I could while preserving the overall integrity of the regular formulae group. When I later extended the term "regular formula" to the other nouns in Homer, I decided to keep the same minimum number, although some of the criteria I elected for choosing 6 as a minimum for the nominative proper nouns (such as always being noun-epithetic) were no longer valid for frequent formulae displayed by nouns in oblique cases.

Since the criteria I originally used for choosing the minimum number were qualitative, it was possible to feel, as long as I was speaking of proper nouns, that the distinction between regular formulae and infrequent formulae was qualitative as well as quantitative; but with the extension of the number to common nouns and the ensuing questionability of some of the criteria (not to mention the fact that none of the criteria had actually been used in determining the minimum number for the common nouns), I seemed to be forced to rely upon intuition to support the distinction for all except the proper nouns. Moreover, some of the mathematical equations about regular formulae that worked splendidly for the proper nouns worked less well for the rest. It was amidst such uncertainty that I encountered the hyperbola depicted on Graph F-O1.

The x-axis reads "one-occurrence-only, two-occurrences-only," and so on. (Let me remind the reader that a formula that is never repeated exactly, only inexactly, is counted as occurring only once.) The y-axis tells us how many instances correspond to each x-point—how many formulae occur that many times. Thus the point ($x = 1$, $y = 673$) represents the fact that 673 different formulae occur just once; the point (2, 490) the fact that

490 formulae occur exactly twice; and so on. The curve describes a smooth hyperbola with a very sharp angle, though there is an interesting flattening at 6-7, then a resumption of the curve. To supplement the graph, I adduce on Table F-O1 the figures for the first 25 points on the x-axis beginning at x = 1.

<u>Graph F-O1: Formulae-occurrences curve, Homer 190 nouns</u>



<u>Table F-O1: Formulae-occurrences, Homer</u>

x:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
y:	673	490	194	98	74	62	60	35	28	20	14	12	10	8	7	2	4	3	7	3	5	9	2	1	2

As the eye moves from left to right on the table and graph, we can follow the sharp downward movement to x = 6, the flattening from x = 6 to 7, another sharp drop from x = 7 to 8, a shallower drop from x = 8 to 11, and then a very gradual drop from x = 11 to 25, with much sporadic up-and-down movement along the way—so sporadic that there are more formulae at x = 22 than at x = 14. The equation for this curve is **y = 736/x – 31.7, R = .97, s = 32.1**.[24] We shall, however, concentrate on the curve

[24] The root-mean-square residual seems very large, but it must be remembered that the y-axis is also very large, reaching y = 673. Despite appearances, the value at x = 2 is the most deviant.

rather than the equation, since it gives a clearer picture of the phenomenon.[25]

It is obvious to the reader that the beginning of the bend in the curve at x = 6 corresponds to the previously selected minimum for regular formulae. Moreover, there is a sharp drop from x = 7 to x = 8, and a very much shallower drop thereafter; 8 was another, almost equally valid choice for a minimum number. The reader may feel that the true leveling-off begins at x = 11; we might also make this our minimum. The graph seems to say that the choice is somewhat arbitrary: the minimum cannot be on the sharply vertical left-hand tail running from x = 1 to x = 5, which must correspond to infrequent formulae, nor on the very gradual right-hand tail, which must correspond to regular formulae; it must lie somewhere on the bend in between, running from x = 6 to x = 11. We might indeed have three classes—infrequent formulae, regular formulae, and transition- formulae—but the gain in precision would probably not be worth the encumbrance to our calculations. Let us be satisfied with the fact that the hyperbola, if we can explain it, endorses our previously chosen minimum of 6 for regular formulae, while indicating that either of the alternatives then available, 8 or 10, would have been acceptable.

If we had plotted Graph F-O1 and had come up with a steadily declining straight line, we could still have made a distinction between regular formulae and infrequent formulae, but it would have remained a quantitative distinction. The existence of the hyperbola suggests that there may well be a qualitative distinction—provided that we can explain the curve's shape. Now not every possible explanation will help us. Consider, indeed, the explanation that seems at first sight the most obvious, that our hyperbola simply follows the pattern of another (conjectural) hyperbola, one that traces the number of times each noun occurs. That is: suppose that we should find that a great many nouns occur exactly 13 times (the number of occurrences per noun that we have chosen for a minimum), a considerably smaller number 14 times, and so on, with a steep descent down to 19 or twenty times, and then a flattening out, so that around 20 to 25 times we have only three or four nouns occurring that often, around 30 to 35 only one or two. This distribution would give us a curve of the same shape as the Graph-F-O1 hyperbola, with number of occurrences per noun on the x-axis, and number of different nouns on the y-axis. We would have

[25] The curves for the *Iliad* and the *Odyssey* are identical to the curve for Homer's 190 nouns, and therefore not worth reproducing. The equations for the individual poems have different parameters from the equation for the 190 nouns, of course, since the numbers on their y-axes are not nearly as large.

a relatively large number of nouns capable of generating formulae that occur once, which would explain why we have a large number of such formulae. We would have a relatively small number of nouns capable of generating formulae that occur more than 14 or 15 times, explaining why their number is so small. We would have a bend in the curve at around 20 times, to explain the sharp change in the slope of our hyperbola at around 6 times on the x-axis. With this as the explanation, we could hardly argue that the bend in our hyperbola is due to a qualitative distinction between regular formulae and infrequent formulae.

Fortunately for our hopes for such a distinction, there is no such number-of-nouns/occurrences-per-noun hyperbola; the conjecture falls apart the moment we look at the figures. They are worth looking at; not only do we rebut an unwelcome hypothesis, but we garner some useful information along the way. We construct a table: one row will read "exactly 13 occurrences, exactly 14 occurrences," and so on, and the other row will tell how many nouns occur that many times. The statistics for the first 25 levels may be found on Table N-O1.

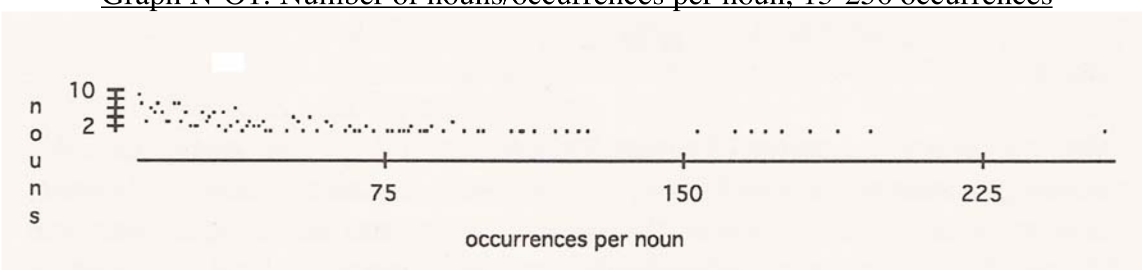
Table N-O1: Nouns-occurrences, Homer

occ/n	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
nouns	9	7	3	6	5	7	5	3	3	7	7	3	5	2	2	2	5	3	4	5	2	5	1	2	6

Nine nouns occur exactly 13 times, 7 nouns occur 14 times, and so on. We observe at once that there are *not* a great many nouns that occur 13 times: there are only 9. It is not true that only three or four nouns occur between 20 and 25 times: there are as many nouns that occur 23 and 24 times as occur 14 times. It is not true that only one or two nouns occur as often as 30 to 35 times: there are 6 that occur 37 times, and 6 that occur 16 times. It is not true that only a small number of nouns are capable of generating a formula that occurs 15 times: a noun that occurs only 23 times (δάκρυ in the *Iliad*) is capable of it, and 135 out of our 190 nouns occur 23 times or more, 54 of them appearing in the above table. If we were to plot the graph for the numbers on Table N-O1, we would get a scattering of points that is, if anything, linear and not hyperbolic. There would be a down- ward movement from level 13 to level 37, but a very gentle one. The graph would certainly bear no resemblance to our hyperbola as a whole— though it would not be dissimilar to its long right-hand tail. And the reader can see this by contrasting the figures on Table N-O1 with the figures on Table F-O1 above: the left side of Table F-O1 has nothing in common with Table N-O1, but the right side matches it very closely.

This very fact may help us in our search for a qualitative distinction between regular formulae and infrequent formulae. If we can show that the shape of the right-hand tail of the hyperbola is indeed caused largely by the frequency of occurrence of the nouns themselves, then the shape of the left-hand tail must have a different cause; or perhaps there are facts about the regular formulae that *enable* them to occur as a function of the occurrences of their nouns, facts that the infrequent formulae lack. Either way, we get a quality or qualities responsible, at least in part, for the distinction. Let us begin by actually plotting on Graph N-O1 the number of nouns vs. occurrences-per-noun, this time including all 190 nouns. We note the dissimilarity between Graph N-O1 and Graph F-O1; if anything, the relationship on Graph N-O1 is linear. If we do give it a linear analysis, we get an equation with a very low slope, minus 0.02, and a root-mean-square residual of 1.6. That is, the number of nouns per level is nearly constant, but there is a slight downward movement as we go from left to right. Naturally enough, since we expect a narrative poem to display fewer nouns that occur 100-150 times than occur 15-20 times.

Graph N-O1: Number of nouns/occurrences per noun, 13-256 occurrences



On the other hand, the resemblance between Graph N-O1 and the hyperbola's right-hand half from x = 11 or 12 on out along the x-axis to x = 79 is striking. After x = 11 on Table F-O1, the numbers go up and down, with a very gradual overall downward movement, a movement often arrested, so that we see, for example, 8 formulae occurring 14 times, and 9 formulae occurring 22 times. This is exactly what is happening on Graph N-O1 and Table N-O1: the numbers go up and down, but there is a gradual and often arrested downhill movement. If we go out far enough on the graphs, and if we extend Table N-O1 to include all the instances on Graph N-O1, we come to a point where for each x value, y = 1. This helps us to understand the gradual decline, as such, in the right-hand tail; it must reflect the equally gradual diminution in the number of nouns that occur often enough to provide a formula that can occur that often.[26] Most nouns occur too seldom to generate a formula exactly repeated 66 or 79 times.[27]

But it does not explain the *sporadic movement* in the right-hand half, the fact that a regular formula is just as likely to occur 22 times as 14 times. Nor does it fully explain why the decline is so gradual, in contrast to the steady, sharp decline on the left, the fact that an infrequent formula is less likely to occur twice than once, three times than twice. A regular formula on the right seems to be enabled to occur as often as the number of overall occurrences of its noun permits; an infrequent formula on the left is obviously prevented from doing the same.

Now the most obvious reason for the difference between the tails is that there are so many more infrequent formulae than regular formulae, 1529 vs 307. The infrequent formulae must be answering poetic needs, each one of which arises rarely, but which as a type arise very commonly. Only once in the *Iliad* does the poet need, or elect, to say in the genitive

---

[26] Naturally we ask whether there is a linear relationship between the number of total occurrences of our 190 nouns and the number of formulae that occur at each level of occurrence (i.e., between Graph N-O1 and the right-hand tail of the hyperbola), and indeed there is. But the correlation coefficient is only moderate high (.51) because there are various factors that work to decrease correlation. See further below.

[27] There is another idea that looms temporarily as a way of explaining the formulae-occurrences hyperbola. It happens that a large number of nouns have formulae that occur once, a significantly smaller number have formulae occurring twice, and so on until we reach 17-22 times, after which the numbers level off and a noun is as apt to have a formula occurring 40 times as occurring 20. Plotting a graph, with x = "possessing a formula occurring x times," and y = the number of nouns that have a formula occurring that often, we get another hyperbola, one also possessing a sharp break, but in a place further out on the x-axis, around x = 18. Moreover, there is a high correlation (coefficient of .94) between the y-values for this hyperbola and the y-values for the formulae-occurrences hyperbola, and an even better one (coefficient of .99) between their logarithms. (That is, the relationship between the y-values is not quite linear, but follows a gentle curve.) It is no doubt true that these two phenomena are closely related; but have we explained anything? This line of argument seems only to postpone the difficulty: *why* do a much larger number of nouns have formulae occurring once than have formulae that occur 18 times? And why such a sharp break around x = 18? Is not the answer going to be, because there is a much larger number of different formulae that occur once than occur 6 times, and there is a sharp break at x = 6? In other words, a great many nouns have formulae occurring once because a great many formulae occur once. The high correlation between the two phenomena means only that each noun tends to have the same *proportion* of formulae occurring once, twice and so on as every other noun. This is an interesting fact, but it does not explain the formulae-occurrences hyperbola.

before position 9 that Achilles was "great of soul," but he says 9 times in this place and case that *someone* was. Just once does Odysseus "rise up," ἀνίστατο in position 6-8, but the verb occurs in this position 7 times in the *Iliad* and twice in the *Odyssey*. The astonishing number of different infrequent formulae, and the fact that there are only 5 places on the x-axis of Graph F-O1 where infrequent formulae can fall, means that even before we constructed the graph, we knew that whatever the left-hand side of the graph might turn out to look like, it could not have been a linear extension of the right-hand tail backwards to the y-axis. It might have been a scattering of points, or a horizontal line, or an S-curve, but whatever shape it took, it had to be much higher on the y-axis than the right-hand tail. But this does not account for the shape that the left-hand tail does in fact take, for its steep and regular decline as opposed to the gradual sporadic decline on the right. It is as if some force were at work upon the infrequent formulae to produce the sharp decline. This force is constrained on the right-hand tail by some counter-force that allows the regular formulae to occur freely. The constraint starts to gain control at x = 6 and is fully in command at x = 11. Indeed we shall see that it begins to exert itself even earlier, and slows down the sharp left-hand decline.

We shall encounter several reasons for the shape of the left-hand tail before we are done, but it is useful to begin with a generality. Entropy is the natural tendency of any system towards maximum randomness. If we recognize formularity—as we have defined it—as a state of order, then with respect to formularity the state of maximum randomness is a non-formulaic occurrence, and entropy is the force at work upon the infrequent formulae. The set of 70 nouns, for example, shows 1204 non-formulaic occurrences. If we subtract from this set those formulaic occurrences, and only those, that are enabled to occur freely by the (as yet unspecified) constraint upon randomness just mentioned, there remain just 747 formulaic occurrences. Without the constraint, there would be many more non-formulaic than formulaic occurrences. Without it, the formularity of our 70 nouns would be 38% instead of the 74% it actually is. Without it, a non-formulaic occurrence has a greater chance to occur than a formula.

The first stage of order is a once-only formula, a partial repetition, which is likelier to occur than a total repetition (if entropy is allowed to play freely). For μεγάθυμος ᾿Αχιλλεύς to count as a once-only formula, μεγάθυμος need not occur again with ᾿Αχιλλεύς, but need only be found with some other noun in the same position; or the two could occur together, but in a variety of different positions. For it to have counted as a twice-only formula, every part of it would have had to be exactly repeated.

Hence—since at this point on the graph entropy is more powerful that the constraint upon it—there are twice as many once-only as twice-only formulae, and a formula is twice as likely to occur once than twice.[28] For a formula to occur three times, the circumstances permitting its occurrence must be present three times, and this is less probable than their presence just twice. (Three of a kind is less common in poker than a pair.) And so on, until we reach the place on the x-axis of Graph F-O1 where the constraint upon randomness is largely in place, somewhere between x = 6 and x = 11.

The best candidate for this constraint has two aspects: the structure imposed by the poets upon the hexameter line to facilitate the use of formulae, viz., the caesurae and the cola, especially the major cola; and the nature of the formulae that are especially devised to fill the major cola. The epic tradition has broken the hexameter line into four segments separated by caesurae. These caesurae can fall in six possible places: after verse-positions 2, 3, 5, 5.5, 7, and 8.[29] Caesurae are determined by cola, and the cola that chiefly operate to restrain randomness are the major cola. Most formulae in the right-hand tail occur in major cola; most infrequent formulae (57%) do not; and the more often a noun-formula occurs, the likelier it is to fall in a major colon: 31% of once-only formulae fall in a major colon, 44% of twice-only formulae, and so on.[30] This growing percentage of occurrences in a major colon slows the steep decline of the left-hand tail, begins to arrest it at x = 6 on Graph F-O1, where most formulae are falling in a major colon, and has brought it to a halt by x = 11, where almost all are.

---

[28] It is twice as easy provided that the poet has the means—largely Hainsworth-alteration and generic epithets and verbs—to create once-only formulae. If he did not, or if they were severely curtailed, the hyperbola proper would start at twice-only. For that reason we might speak of these means as another kind of constraint upon randomness. See further below.

[29] I am accepting the formulation of Geoffrey Kirk, Berkley Peabody, John Foley, and others: see Foley 1990:73-84.

[30] In any discussion of the major cola, we must use the figures for the 70 nouns in Homer, since the calculation for the 190 nouns is not yet complete. Complete figures for these 70 will be found below; figures for nominative noun-epithet formulae for the 38 characters who occur more than 20 times may be found in Sale 1989:387-88. On the basis of these samples, and of non-statistical examination of all 190 sets, we can say that the statements in the text are certainly true for all proper nouns in the nominative and for a representative sample of all common nouns; they are *almost* certainly true for all nouns in all grammatical cases.

To fill these major cola, the tradition devised the formula-systems, many of which are so elegantly isolated and analyzed by Milman Parry. We have regular verb-formulae that put the verb in the identical position each time, admit of relatively few variations in the words accompanying the verb, say only a few things ("spoke, perceived, rejoiced, obeyed, smiled, departed" cover almost all of their semantic range), and usually occupy just four positions (up to the trochaic caesura, up to the hephthemimeral caesura, up to the bucolic diaeresis, and 1-2...9-12). Matching these verb formulae is a much larger number of regular noun-epithet formulae filling out the remainder of the line. When one of these complementary pairs can be used, randomness (from the point of view of meter and formula) is virtually eliminated. And even when a regular noun-epithet formula must be employed without a matching verb-formula, it inevitably reduces the number of syntactic, metrical and semantic possibilities available to the rest of the line, and thereby imposes a certain amount of order upon it. The common-noun regular formulae also display a few noun-verb formulae that operate in a similar fashion.

Twice as many major cola fall in the second half of the line as fall in the first. The principle of major cola as a constraint on randomness thus dovetails with what John Foley calls "right-justification," the overall tendency for the hexameter line to display greater phraseological and metrical fixity in its second half (Foley 1990:56-57, following Roman Jakobson, Gregory Nagy and others). At first sight, the regular verb-formulae we have been discussing seem to challenge this tendency, since they display fixity, in that they are exactly repeated, and they fall at the beginning of the line. But there are not very many such different regular verb-formulae (I count just 11 that reach the trochaic or hephthemimeral caesura), while the number of different regular noun-epithet formulae that can be used to match them is very large. Or used, indeed, for other purposes: among our 190 nouns there are 37 proper nouns in the nominative case, with 1178 regular formulae-occurrences, and only 330 of these occurrences, by a preliminary count, match verbal regular formulae that open the line. As a result, noun-epithet regular formulae falling in second-half major cola are often found matched with line-openings that are not regular formulae, or not formulaic at all by the definition of "formula" that I am using, and hence more free, less constraining of randomness. Thus the primary source of constraint comes from the noun-epithetic major cola, and these mostly fall in the right-hand portion of the verse.

Constraint can therefore be seen as arising from the colonic system as such, with its ubiquitous major cola, and from noun-epithetic regular formulae, often supplemented by verbal regular formulae. The colonic system has created a ubiquitous need, the need to fill a major colon. The noun-epithetic regular formulae come into existence to help meet this need. They often require a matching verbal regular formula, but when they do not, they still demand to be complemented syntactically and semantically, and this in itself serves as a constraint—less particular, to be sure, but not negligible. In short, the colonic system constrains the line of verse to accommodate the regular formulae; if a regular formula is appropriate, that regular formula meets the need to fill the major colon, and thereby imposes its own demands upon the rest of the line and indeed the context generally. When this demand is for a verbal regular formula, the whole line is mostly determined; when it is not, randomness is increased, but within limits.[31]

How does the constraint help determine the shape of the right-hand tail? Clearly—since not every major colon contains a regular formula—the major cola are not so much causative as enabling; they obviate the effect of randomness, but do not determine exactly how often a regular formula will appear. The frequency of occurrence of a regular formula is actually caused by five other factors: the number of times the noun itself occurs, the localization of the noun, the syntax and meaning of the regular formula, the ability of the regular formula to extend itself into other cola, and the existence of other regular formula for the noun. The phrase δῖος ᾿Οδυσσεύς in the *Odyssey* occurs 79 times, the largest number for any noun-formula. It owes this frequency in part to the fact that ᾿Οδυσσεύς occurs more often than any other noun, 256 times.[32] It owes it to the fact that the word is highly localized, almost always occurring in final position; it does not stray into other parts of the line, where the regular formula is unusable and infrequent formulae must be employed. It owes it to being noun-epithetic, and to the epithet's being context-free: the formula can be

31 Two qualifications: it goes without saying that the semantic and aesthetic needs of Homer are far too various to be satisfiable inevitably by a regular formula; but the need to fill a major colon with something remains nearly perpetually. And verbal regular formulae are not fully determined; they include metrically identical alternatives, and some include participles that can be replaced as the context requires.

32 It is important to stress that these and comparable totals include no alternate names or spellings (such as ᾿Οδυσεύς) and no other grammatical cases.

used anywhere in the poem.[33] It owes it to the fact that the formula is commonly extended backwards to the trochaic caesura with the additional context-free epithet πολύτλας. And it owes it to the fact that there is only one other regular formula for the noun, πολύμητις ʼΟδυσσεύς.

If the first of these factors were the only one, we would expect close correlation between sporadic movement of the right-hand tail of Graph F-O1 and the similar movement on Graph N-O1—that is, between the numbers on the right of Table F-O1 and the numbers in Table N-O1. Now there *is* a correlation, but the coefficient is only moderately high (.51). Correlation has been reduced by the play of the other four factors we have just enumerated. It has been reduced by the tendency for a noun with a lower localization to display a lower percentage of regular-formula-occurrences.[34] Indeed, when we examine the 100 nouns that generate regular formulae occurring just 6 or 7 times, we find that many of these occur very frequently, and that many (70%) also have low localization; thus it will often be low localization, not a low total of the noun's occurrences, that is responsible for the existence of the infrequently occurring regular formulae. Correlation has been reduced by the presence of noun-verb regular formulae, which can only be used when the action that they refer to happens; no noun-verb formula occurs more than 13 times. It has been reduced by the presence of formulae that cannot easily extend themselves backwards, a frequent phenomenon with common nouns, which (unlike proper nouns) are almost never extended by adding one adjective to another; they extend, if they do, with verbs instead, which are less free of context and therefore cannot be used nearly as often. And it has been reduced by the presence of other regular formulae for the noun, developed in part because of limitations on extension, and in part because nouns with low localization sometimes form regular formulae while occupying an unusual position (νῆας in the *Iliad* forms regular formulae in four different

---

[33] Most regular-formula-epithets are context-free: Diomedes' war-cry is always splendid, whether he is shouting or not; Achilles' feet are swift even when he is asleep; and so on. So if the poet needs to say "Diomedes," Diomedes' regular formulae will almost never say the wrong thing, and by epic convention will therefore almost always say the right thing. See further the discussion in Sale 1989:389-90.

[34] See Sale 1989:372-77, 410. The correlation between localization and percentage of regular-formula-occurrences for the 22 frequently occurring characters discussed there has a very high coefficient, .92. The coefficient is lower, .71, when we add the rest of the proper nouns and all the grammatical cases, and still lower, .58, when we include common nouns; but even the last figure points to a genuine relationship.

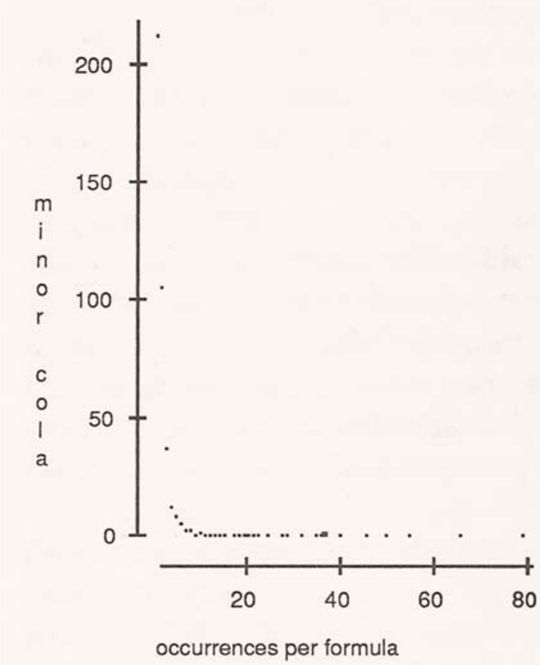
positions, and consequently displays 7 different regular formulae).

To sum up this portion of the argument, the major cola permit the free occurrence of formulae of a certain kind and shape, and this explains the low *slope* of the hyperbola's right-hand tail: it imitates the slope of the nouns-occurrences graph N-O1. Meanwhile, the irregular movement on graph N-O1, together with the other four factors just discussed, explain the sporadic up-and-down movement on the hyperbola, why there should be more regular formulae occurring 22 times than 14. The other four factors tend to lower the number of times a regular formula will appear—or rather, to raise the number of less-frequently-occurring regular formulae. The first factor, in contrast, will tend to spread the regular formulae out along the x-axis: there are 6 nouns, for instance, that occur 37 times and 6 that occur 16 times. Eventually, of course, the supply runs out, and only one occurs 197 times, one 256 times, and none in between and none after that.

We have therefore explained the very uneven and equally gradual decline of the right-hand tail, and can return to the left. We have already said earlier that it exists in part because there are so many infrequent formulae, five times the number of regular formulae, and that these infrequent formulae must exist to meet needs that rarely arise for any given noun, but are of a sort that arise frequently. We have also argued that its shape is due in part to the struggle against entropy, to the fact that without the presence—or rather with the considerably diminished presence—of the constraint that supports the free occurrence of the regular formulae, it is more difficult to have a formula than a non-formulaic occurrence, more difficult for a formula to occur twice than once, three times than twice, and so on. Just as it is the ubiquity of the constraint that causes the very low slope of the right-hand tail, so its reduced presence causes the steep slope on the left.

We have partly explained the left-hand tail, but we are faced with some bewildering questions. Why are the constraints not always in place? Why do we have this vast horde of infrequent formulae? Why do the regular formulae not do the job? If the constraints were always in place, would we get a linear curve on Graph F-O1? And why are there relatively few different regular formulae? It would be interesting to attempt an answer to each of these questions, but to save time here I suggest that we look at the job that the infrequent formulae do in fact do, and see whether this might not explain, at least intuitively, why they exist, and in such large number. (I shall do this in detail in Appendix 1; here let us summarize.) We have already identified one of their tasks: infrequent formulae answer to rare metrical needs by filling in minor cola. Though it is true that almost all lines of hexameter verse include a major colon, a good many lines also include rarer ones, cola that a given noun is not likely to occupy more than a few times in the course of the poem. If a noun in this position is embodied in a formula that fills the rare colon, that formula will usually be an infrequent formula. True, it will happen that some nouns do occupy a rare colon more than a few times; there are some regular formulae that fall in minor cola, but not many. Based upon our sample of 70 Homeric nouns, while 57.5% of the infrequent formulae do not occupy major cola, only 8% of the regular formulae do not, and none of these 8% occurs more than 10 times. The existence of these rarer cola obviously adds variety to the line of verse; such variety is built into the Homeric technique, which is much more flexible in this respect than the technique of the *Chanson de Roland*.

Graph F-O2: Minor-colon formulae, Homer



Graph F-O3: Major-colon formulae, Homer

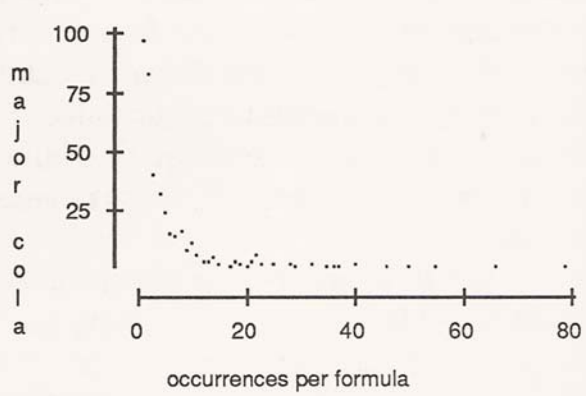


Table F-O2: Minor-colon formulae, Homer

x:	1	2	3	4	5	6	7	8	9	10	11etc
y:	216	106	38	14	6	9	3	4	0	1	0
%:	70	56	50	30	20	43	19	22	0	8	0

Table F-O3: Major-colon formulae, Homer

x:	1	2	3	4	5	6	7	8	9	10	11
y:	94	84	38	32	24	12	13	14	8	11	6
%:	30	44	50	70	80	57	81	78	100	92	100

If we separate all the formulae that fall in major cola from those that do not, and plot formulae-occurrences graphs, we get the picture given in Graphs F-O2 and F-O3 and Tables F-O2 and F-O3. Keep in mind that these figures are for 70 nouns only (35 proper nouns in the nominative plus 35 common nouns), since the study of the major cola for the 190 nouns is not yet complete. The graphs were made unequal in size in order to maintain the same scale and bring out the fact that the decline in numbers is considerably steeper for the formulae that do not occupy major cola.

Note that the percentage of occurrences in major cola rises steadily from 31% until it reaches 100% when y = 0 and x = 9.[35] Of course both shapes bear a resemblance to the shape of the Graph-F-O1 hyperbola: a sharply decreasing left-hand tail, and a long flat right-hand tail. The right-hand tail of Graph F-O2 corresponds to the fact that almost all regular formulae fall in major cola: after x = 9, y almost always = 0.[36] The right-hand tail of Graph F-O3 has the familiar low slope and irregular descent. The left-hand tails continue to indicate that there is a huge number of infrequent formulae, and that the number declines rapidly as the number of occurrences per formula goes up. But the difference between the left-hand tails of the two graphs means that whatever is causing the decline on the left may not be working at all on the right.

This cause, of course, is entropy; and since we have seen from discussing the right-hand tail of our original hyperbola on Graph F-O1 that the major cola effectively constrain entropy, we must conclude that entropy

[35] The percentage of occurrences in major cola reaches 100 at x = 9, then dips at x = 10 to accomodate just one formula, ’Αντίνοος προσ(μετ)έφη Εὐπείθεος υἱός. I cite it in its extended form, with the verb; the alternation προσ/μετ is a function of whether the addressee (referred to at the beginning of the line) is singular or plural, and it could be argued that even in the full form we have 10 examples of the same formula. This tempts us to try to argue that the formula fills the major colon from 3 to 8. But it cannot, and does not, exist without the final word, and therefore it is simply too long for the colon. I cite it to show how close we are to being able to say that for these 70 nouns all formulae occurring more than 8 times fill a major colon.

[36] If we had all 190 nouns, we would get a few more non-major-colon formulae, but only a few; three of them are discussed in note 11.

is indeed not at work on Graph F-O3. And yet it still has a left-hand tail! Let us continue to examine the tasks that infrequent formulae perform, to see why we continue to have so many infrequent formulae even after those falling in minor cola have been subtracted.

One task is to be a *noun-verb* formula. Most noun-verb formulae are infrequent formulae: out of 382 major-colon formulae (regular formulae and infrequent formulae) for the 70 selected nouns, 164 are noun-verbal, and 151 of these, 92%, are infrequent formulae. If we were to plot a formulae-occurrences graph for major-colon noun-verbs, the shape would be similar to Graph F-O3, only still less steep, and also less regular, since there are more major-colon noun-verb formulae occurring 4 times than 3. Why, then, are noun-verb formulae mostly infrequent formulae, so that we continue to see a left-hand tail?

There are at least four reasons, the first a negative metrical consideration. In order to repeat a major-colon noun-epithet formula with a context-free epithet, virtually the only thing necessary is that the person or thing referred to by the noun occur twice in the text. The interplay between noun-epithets and the major cola created by the regular formulae is such that it is extremely easy for the poet to create a line with a major colon for a noun-epithet formula to fill; the model for the rest of the sentence is already there. A noun-verb formula, on the other hand, must be fitted into a sentence that is less easily made ready for it, so that an alternative way to say what is wanted may be chosen the second time the idea is expressed.

Second, a noun-verb formula has a complex referent. The fact that so many noun-epithetic formulae contain context-free epithets means that, practically speaking, the referent of the noun is *the* referent; the epithet has no limiting role. There is almost no such thing as a context-free verb in the epic vocabulary; to use a noun-verb formula the poet must want to mention not only a particular person or thing, but also a particular action or state of affairs. No character referred to in the nominative, not just among our 70 nouns but in all of Homer, possesses a noun-verb regular formula, apart from the extension of ῥοδοδάκτυλος Ἠώς by means of φάνη. Even the common nouns among our 70 display only 13 such regular formulae, as opposed to 151 infrequent formulae. Hands are frequently raised (in the *Iliad*), or laid upon food (in the *Odyssey*), ships frequently arrived at (in the *Iliad*), κῦδος given and won (in the *Iliad*), but most things happen more seldom. And in fact a good many actions are performed just once, though they may belong to a *class* of actions that occur more often. The phrase Αἴας δὲ κορύσσετο, for instance, which occurs once, is a formula because the verb is generic, the same verb-form being used also in the same position of Patroclus and Achilles. We can imagine the phrase's having occurred twice; but it is unthinkable that it should be a regular formula (the verb is used only once even for Achilles, who is mentioned much more often than Ajax). The act of arming occurs relatively often; but Ajax simply does not, and in any remotely similar poem will not, arm himself more than once or twice. Thus a basic task of major-colon noun-verb infrequent formulae is to say things that rarely need to be said.[37]

The third cause applies as well to major-colon noun-epithets. Some major cola are much more rarely occupied by noun-formulae than others. For the selection of nominative proper names that he singled out for special study, Parry identified 4 major cola: 1-5 (up to the penthemimeral caesura), 6-12 (after the trochaic caesura), 8-12 (after the hephthemimeral), and 9-12 (after the bucolic diaeresis). For oblique cases and common nouns he added 1-5.5 (up to the trochaic caesura) and 5.5-12 (after the penthemimeral caesura); he recognized that certain nouns with a rarer but normal metrical shape fell in rarer cola. I in turn have added 2-8 and 3-8, moved by exactly the same considerations: certain nouns, especially certain proper nouns in the nominative that Parry did not single out, put their frequent formulae in these cola; and indeed certain complimentary verb formulae are shaped to fit around them, at 1-1.5 (or 1-2) plus 9-12 (τὴν δ' αὖ. . .ἀντίον ηὔδα, for instance). But there are 3 cola that noun-formulae occupy far more commonly than the others: 5.5-12, 8-12, and 9-12.[38] Hence when we observe 18 once-only major-colon noun-verb formulae falling in the rarer major cola, and only one of them occupying the colon where its regular formula falls, we conclude that at least the other 17 owe the scarcity of their occurrences to the rarity (relative to the meter of their nouns) of the cola they occupy. Only 2 of these 17 put the noun at its

---

[37] It might be objected that just because a verb occurs rarely, the idea need not occur rarely. But I have not noticed any instances where two different verbs used with a given noun in the *same major colon* say the same thing. If it does happen, it happens very seldom; always, or almost always, the need is as infrequent as the infrequent formula that meets it. Of course some needs are similar to each other. There are two noun-verbal infrequent formulae, for instance, that occupy the same major colon as χεῖρας ἀνάσχων (a regular formula in the *Iliad*) and mean something akin: χεῖρας ὀρεγνύς, and χεῖρας ἴαλλον (an infrequent formula in the *Iliad*). But ἀνάσχων is an action appropriately directed towards gods alone; towards mortals we use ὀρεγνύς, a different action, while for food we use the formula χεῖρας ἴαλλον .

[38] That 1-5 is much less common—one-tenth as frequent—as each of the others is clear from Parry's own figures (1971:39, Table 1).

localization-point, making it all the more reasonable that the remaining 15 should occur only once. In sum, another basic task of noun-verb infrequent formulae is to occupy cola that their nouns, and indeed most nouns, rarely occupy, and therefore to provide formulae for these nouns when they are wandering away from their localization-points.

There is a fourth cause, and that is accident. I have noted 5 major-colon noun-verb formulae that could have been regular formulae; three of them are, in fact, regular formulae in the other poem. Note that these formulae are still meeting rare needs; there is no reason why a need cannot be accidentally rare. A poet can easily happen to mention a person, an object, or an action less frequently in one poem than he might have in another. What is astonishing is that as few as 5 noun-verb formulae are infrequent for this reason.

We can now subtract the 164 major-colon noun-verb formulae from the 382 major-colon total formulae, and construct a formulae-occurrences graph for the remaining 218 noun-epithets. It too has a left-hand tail, but much shallower. There is a difference of only 2 between the 39 that occur once and the 37 that occur twice; then comes a steeper falling off, and then the graph grows level and begins the right-hand tail, the very gradual descent, at x = 4, y = 12. There are 121 infrequent formulae and 97 regular formulae. Again we ask what job it is that the infrequent formulae, this time noun-epithetic infrequent formulae, perform such that they are infrequent formulae and the graph continues to possess a left-hand tail. The answer becomes more complex, and we shall look at it in greater detail in Appendix 1; let us merely sketch it here.

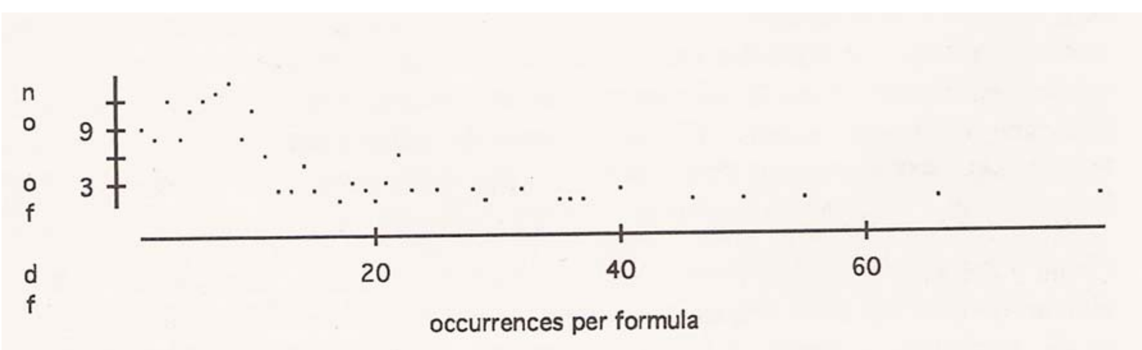
Of the 121 major-colon noun-epithetic infrequent formulae, a total of 35 meet rare *metrical* needs. Some 23 of these occupy the rarer major cola: again the *sort* of need is common, but the rarity of the need for the individual formulae is underscored by the fact that each of the 23 occurs only once or twice. There are 12 more that offer rarely needed metrical alternatives to other formulae, usually regular formulae, falling in the common major cola (the infrequent formulae can begin with a double consonant, for instance).

We also have rare needs of a *semantic* or *aesthetic* nature: 28 of the major-colon noun-epithetic formulae are specific to the context in which they appear, and 12 seem to be used for special effect. The phrase Θέτις κατὰ δάκρυ χέουσα is a good example of the former, since it can only be used if Thetis is weeping. The phrase μεγάθυμος Ἀχιλλεύς is an instance of the latter; the poet wanted to avoid πόδας ὠκὺς Ἀχιλλεύς, which would have ineptly anticipated πόδας in the next line.[39] Obviously these special-effect formulae are not likely to occur more than once or twice; only one out of 12 occurs more often. There are thus 40 formulae that meet, or probably meet, semantically or aesthetically determined rare needs.

There are 3 formulae whose is existence is something of a puzzle. They not only overlap other formulae metrically, but they appear to be genuine violations of the principle of economy in that they are hard to defend as semantic or aesthetic alternatives to the formulae they overlap (see Appendix 1).

That leaves 43 major-colon noun-epithetic formulae that appear to be infrequent formulae by accident—formulae that could be regular formulae, or indeed are regular formulae when used in the other poem. They fall either into the commonest cola—into cola that are frequently occupied by nouns with their meters—or into cola where they put the noun in a frequently occupied position; they have context-free epithets; they are not aimed at a particular metrical, semantic, or syntactic effect; and they do not perform the same job as an already existing regular formulae. We have also already noted 5 noun-verb infrequent formulae that have this appearance; only 5, since we have seen that the bulk of such formulae could never be used more than a few times. There are thus 48 formulae that might well have been regular formulae under different circumstances—if, for instance, the poet had been using these nouns more often, or if certain metrical circumstances had happened to arise more often. Let us combine them with the 110 major-colon formulae that actually are regular formulae, and plot their formulae-occurrences curve on Graph F-O4.

Graph F-O4: Major-colon **rf** and accidental **if** vs. occurrences-per-formula



[39] I owe this example to Hainsworth 1968:9, n. 2.

At last the left-hand tail has vanished. Though it is theoretically possible to interpret Graph F-O4 as a hyperbola, the fact that the left-hand side is a scatter, not a tail, makes such an interpretation exceedingly unattractive. It makes much better sense to treat the whole as roughly linear with a gently declining slope. Indeed it resembles Graph N-O1, which relates the number of nouns to occurrences-per-noun, very closely indeed. The resemblance is so marked that we are fully justified in attributing the gentle decline on Graph FO4 to the growing lack (as we proceed outward along the x-axis) of nouns that occur often enough to produce formulae that occur that often.

What we have done, therefore, is to subtract from the total number of infrequent formulae all those formulae that clearly answer to rare needs. These needs have proved to be: for formulae in minor cola, for noun-verb formulae, for formulae filling rare major cola, for rarely-needed metrical alternatives, for expressing a meaning specific to a context, and for creating an unusual special effect. By subtracting these infrequent formulae, we have subtracted the left-hand tail from the hyperbola. We have left behind a sporadically descending, roughly linear curve describing the behavior of a group of formulae that have the same characteristics whether they occur once or 79 times.[40] The difference between these characteristics and the rare needs just enumerated gives us the qualitative differences we were seeking between regular formulae and infrequent formulae.

These qualitative differences account, therefore, for both tails of the hyperbola on Graph F-O1. In between the tails is the transitional area, the bend from x = 6 to x = 11, to remind us that there is no real minimum

[40] The graph omits 18 non-major-cola regular formulae and any non-major-cola infrequent formulae that have the characteristics of regular formulae. Since such regular formulae are exceptional, the task of determining what infrequent formulae resemble these exceptions enough to be called "accidental infrequent formulae" is a difficult one. Indeed many of the 18 regular formulae look very much like regular formulae by accident: μέσον σάκος and μένος μέγα, for instance. Remember too that we chose the lowest possible minimum for regular formulae; if we had chosen a slightly higher one, 8, only 6 would remain. On the other hand, there is every reason to expect regular formulae by accident; in the course of a long poem, certain phrases that might be expected to occur rarely will naturally occur a little more often. I might have produced a graph virtually identical with Graph F-O4 simply by removing examples such as these. Rather than winnow both the regular formulae and the infrequent formulae with insufficient confidence in the objectivity of the procedure, I preferred to set the problem aside by basing the graph on the characteristics of the vast majority (86%) of the regular formulae. If one simply includes the other 18 regular formulae, what results is a graph very similar to Graph F-O4 with a greater bulge in the left-center.

number for regular formulae, only a minimum range of numbers. Between 6 and 11 some formulae are in principle regular formulae, others are infrequent formulae that happen to have occurred a little more often, and others are no doubt indeterminate. It may well be that mathematical sophistication will one day enable us to dispense with a minimum number, but for now the interests of statistical simplicity demand that we make a choice, and the hyperbola certainly permits, nay encourages, the choice of 6 (introducing the first flattening), 8 (after the last large drop), or even 11 (introducing the second flattening). In choosing 6 we are electing, for better or worse, to make the regular formula group as large as possible, and therefore, when possession of a regular formula is a criterion for including a noun in a group, making that group as large as possible.

The qualitative differences, then, account for the hyperbola, and the hyperbola, in turn, gives a quantitative picture of the formulaic behavior of Homer's nouns: a small number of frequently employed formulae are used to meet common needs, while a large number of formulae, each one of which is infrequently employed, meet rare needs of a sort that commonly arise. Now that we have given this thorough empirical explanation, it is proper to add that the hyperbola was pretty well predictable on theoretical grounds. The formularity equation, Equation 1A, guarantees that most noun-occurrences are formulaic, and that when **total occurrences** is high, **formulaic occurrences** is high, so that either the number of **different formulae**, or **occurrences per formula**, will be high as well. Equation 2A asserts that when **total occurrences** and **formulaic occurrences** go up, it is primarily not **occurrences per formula** but **different formulae** that goes up with them. Now if **different formulae** were stationary with **total occurrences**, and **occurrences per formula** went up and down, we would probably not have a hyperbola. We would expect each of the formulae of a frequently occurring noun to occur more often, so that such a noun would have few, or no, infrequent formulae, and the left-hand side of the graph would be not much bigger than the right. Given that the reality is the opposite to this scenario, that **occurrences per formula** is nearly stationary with **total occurrences** while **different formulae** goes up and down, we are assured the existence of a large number of infrequent formulae, and entropy will shape most of these into a left-hand tail. Then the fact that there is a cap on the number of regular formulae ensures that, except on the bizarre chance that no regular formulae occur more than 7 or 8 times, we will have a low right-hand tail.

Washington University

References

Bakker 1988 Egbert Bakker. *Linguistics and Formulas in Homer*. Amsterdam and Philadelphia: Benjamins.

Duggan 1973 Joseph J. Duggan. *The Song of Roland: Formulaic Style and Poetic Craft*. Berkeley, Los Angeles, and London: University of California Press.

Duggan 1975 ______. "Formulaic Diction in the *Cantar de Mio Cid* and the Old French Epic." In *Oral Literature: Seven Essays*. Ed. By Joseph J. Duggan. Edinburgh and New York: Scottish Academic Press and Barnes and Noble. pp. 74-83.

Foley 1988 John Miles Foley. *The Theory of Oral Composition: History and Methodology*. Bloomington and London: Indiana University Press.

Foley 1990 ______. *Traditional Oral Epic*. Berkeley, Los Angeles, and London: University of California Press

Hainsworth 1968 J.B. Hainsworth. *The Flexibility of the Homeric Formula*. Oxford: Clarendon Press.

Janko 1981 Richard Janko. "Equivalent Formulae in the Greek Epos." *Menmosyne,* 34:251-64.

Janko 1982 ______. *Homer, Hesiod, and the Hymns: Diachronic Development in Epic Diction*. Cambridge: Cambridge University Press.

Janko 1990 ______. "The *Iliad* and its Editors: Dictation and Redaction." *Classical Antiquity,* 9:326-34.

Janko 1992 ______. *The Iliad: A Commentary*. *Volume IV: Books 13-16*. Cambridge: Cambridge University Press.

Lord 1960 Albert Lord. *The Singer of Tales*. Cambridge, MA: Harvard University Press.

O'Neill 1942 Eugene O'Neill, Jr. "The Localization of Metrical Word-Types in the Greek Hexameter." *Yale Classical Studies*, 8:105-78.

Page 1955 Denys Page. *History and the Homeric Iliad*. Oxford: Clarendon Press.

Parry 1971 Adam Parry, ed. *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Clarendon Press.

Sale 1984 W.M. Sale. "Homeric Olympus and its Formulae." *American Journal of Philology,* 105:1-28.

Sale 1987 ______. "The Formularity of the Place-Phrases in the *Iliad*." *Transactions of the American Philological Association*, 117:21-50.

Sale 1989 ______. "The Trojans, Statistics, and Milman Parry." *Greek, Roman and Byzantine Studies*, 30:341-410.

Visser 1988 E. Visser. "Formulae or Single Words?" *Würzburger Jahrbücher*, 14:21-37.

Appendix I: The Birth of Infrequent Formulae in Homer[41]

The purpose of this appendix is to fill out the picture we have just painted of the various materials the oral poet had on hand, and the techniques in which he was trained, that enabled him to create or employ infrequent formulae in the course of composition. In addition to his regular store Homer possessed, before he composed, a generic store, a store of patronymics, and probably but not necessarily some distinctive formulae that lent themselves to Hainsworth-alteration—a precompositional distinctive store. He had been trained in the process of alteration and in the process of creating a compositional distinctive store as he composed. He had learned how to meet rare needs by creating infrequent formulae with these processes and materials.

As a result, our 70 nouns display a total of 652 infrequent formulae, of which 380 fall in a minor colon and 272 in a major colon. Those

[41] Let me stress once more that all that follows is little more than a detailed examination, mathematically oriented, of the account of composition in performance given in Lord 1960:37-67. Also, the process I shall be describing whereby infrequent formulae are born has a great deal in common with the descriptions given by Visser 1988:21-37, and Bakker 1988:151-95.

infrequent formulae that fall in *minor cola*, though they sometimes put the noun at the localization-point, most often supply a formula for a noun that has wandered into an unusual position in the verse. Such formulae are therefore almost certain to be answering rare needs; the need for a particular formula in an unusual position is necessarily rare, though the general need for a formula of this type is very common. It is often met with a generic epithet or verb, or a patronymic. Ἕκτωρ. . .Διὶ φίλος is found in positions 2 through 8 three times; Ἕκτωρ. . .Βοὴν ἀγαθός in 1 through 9 just once; Ἕκτωρ μεγάθυμος in 2 though 5.5 just once; Ἕκτωρ. . .ἀκόντισε δουρὶ φαεινῷ spread over the whole line occurs 3 times. Another very common device is to place a regular formula in a different position in the line: δολιχόσκιον ἔγχος after the penthemimeral caesura, for instance, instead of after the hephthemimeral; or to alter it further, by inversion (Ἀπόλλων Φοῖβος), separation (κορυθαίολος ἠγάγεθ' Ἕκτωρ), inflection, and so on. Or an infrequent formula can be similarly altered: Αἴας. . .φέρεν σάκος running from 2-3 and 5.5-8 is an instance. Or the poet may combine generic with distinctive: Αἴας δ' ἐγγύθεν ἦλθε φέρων σάκος ἠύτε πύργον, or give something quite distinctive: Αἴας. . .πελώριος ἕρκος 'Αχαιῶν. (Naturally there can be almost no distinctive phrases—phrases that neither contain a generic or patronymic nor are Hainsworth-alterations—that occur just once: we would not be able to identify them as formulae.)

The motive for using such minor-colon infrequent *noun-epithet* formulae is probably primarily (not exclusively) metrical, since most of the epithets are like φαίδιμος, fine and colorful but not highly specific to the context. The poet has decided to let the noun wander, and must accomodate it by filling an unusual colon. But a few have distinctive epithets, such as πελώριος ἕρκος 'Αχαιῶν, and these tend to add real power. The motive behind the *noun-verb* formulae, on the other hand is—as it almost always is––primarily semantic, the need to refer to an unusual action or state of affairs. There are a number of generic verbs that exist for this purpose (ἀνίστατο, κορύσσετο, ἐκέκλετο, to name just a few) but a great many of these formulae are distinctive.

In the relatively infrequent case where a formula in a minor colon puts the noun at the localization-point, we can of course no longer speak of accommodating the wanderer. The motive for noun-verb formulae of this sort is again semantic, to say something that, as with most noun-verb formulae (see below), is not often said; these formulae tend to occupy a whole line, or else to be found in enjambement. Noun-epithet formulae of this sort, on the other hand, are most often alterations of regular formulae (sometimes infrequent formulae), or else patronymic or generic formulae, that put the epithet in an unusual position: κορυθαίολος. . . Ἕκτωρ, Κρονίδης. . .Ζεύς. The motive for most of these appears to be metrical; and it is clear that the need is unusual, not merely because of the position of the epithet but because of the unusual colon occupied.

Of the 272 *major-colon* formulae, 151 are noun-verbal and 121 are noun-epithetic. We had already observed that most of the 164 major-colon total (regular and infrequent) *noun-verb* formulae were infrequent formulae and answered rare needs, and there is little useful to add here. A noun-verb formula is prima facie more likely to be uncommon than a noun-epithet, since the latter has but one referent—the person, thing, concept, and so on that it means—while the noun-verb formula has two, the person and the action. You will simply mention someone far more often than you will say that he or she is engaged in a particular deed. As a result, there are only 5 noun-verb formulae that do not answer rare needs. These I classify as "accidental infrequent formulae"; they might have been regular formulae in another poem. Three of these are, indeed, regular formulae in the other poem and need not be cited here; the other two occur 5 times each and might well have occurred a sixth: Ἴλιον ἐκπέρσαντ(α) and μήδετο ἔργα (the subject of the latter is Zeus). The 121 *noun-epithetic* infrequent formulae, however, need some additional discussion to show which ones answer a need that really is rare, and which should be classified accidental infrequent formulae. First, the 35 that answer rare needs of a *metrical* nature.

We have 23 that occupy the *rarer* major cola, and are obviously meeting rare needs. Five of these (μένεα πνείοντες Ἀχαιοί, for example) are nominative noun-epithets after the penthemimeral caesura; only oblique cases and common nouns have regular formulae here. One, Ἀτρείδης Ἀγαμέμνων in position 3-8, pulls the noun out of the localization point, which it almost always occupies. Six (Αἴας διογενής, for instance) place a spondee or a trochee in 1-5; the regular formulae almost always put such words at the end of the line. Six (such as ἀργυρόπεζα Θέτις in 1-5) are Hainsworth-alterations putting the formula in an unusual place. Two (Τηλέμαχός θ' ἥρως, for instance) create a formula in 1-5 for a choriambic (first paean) noun; such nouns form regular formulae in 3-8. And three have generics that are never used to form regular formulae: ἀντίθεοι Μνηστῆρες, τεύχεα μαρμαίροντα, Ἴλιον αἰπεινήν.

There are 12 more that occupy *common* major cola, where the rarity of the need is slightly less visible. All are metrical alternatives (e.g. initial vowel instead of consonant) to other formulae: 9 to regular formulae, 3 to accidental infrequent formulae. Of the 9 alternatives to regular formulae, 5 begin with a double-consonant used to make position (πτολίπορθος ᾿Οδυσσεύς in both poems, for instance); there are no cases known to me where a noun has 2 regular formulae in the same colon and one has a double-consonant of this sort, and so I consider the need met by the infrequent formula to be rare.[42] We have an example of the reverse situation among the other 4 infrequent formulae that alternate with regular formulae: Μενέλαος ἀμύμων, where a generic epithet produces a single-consonant alternative to a regular formula beginning with a double consonant, ξανθὸς Μενέλαος. Of the remaining 3, two (ὑπέρθυμος Διομήδης and ᾿Αλαλκομενηὶς ᾿Αθήνη) fall in 6-12; for proper nouns in this colon the poet usually prefers a regular formula that will extend backwards from a shorter colon later in the line, with generics (ἄναξ, γέρων, Θεά, μέγας) or not (ποδάρκης, πολύτλας, βοήν). We might cite a situation which could provide us with an analogy for declaring these two to be accidental infrequent formulae: ἐυκνήμιδες ᾿Αχαιοί and κάρη κομόωντες ᾿Αχαιοί are both regular formulae. What keeps me from making this declaration is first, that the generic ὑπέρθυμος is never found in a regular formula, and second, that ᾿Αλαλκομενηὶς ᾿Αθήνη only occurs twice, both times in combination with ῞Ηρη τ᾿ ᾿Αργειή. It seems needed only in this unusual circumstance. The ninth and last alternative to a regular formula is νώροπι χαλκῶι, used when the meaning "armor" is intended and when an initial consonant is needed in place of αἴθοπι χαλκῶι. The specialized nature of this need made me reluctant to classify the formula as an accidental infrequent formula.

The three common-cola metrical alternatives to *infrequent* formulae are ἥρωες ᾿Αχαιοί (beginning with a vowel) and μεγάθυμοι ᾿Αχαιοί (beginning with a double consonant), both alternative to ϝελικῶπες ᾿Αχαιοί (beginning with a single consonant) in 8-12, and εὐήνορα οἶνον, alternative to μελιηδέα οἶνον. The reasons why the need for μεγάθυμοι ᾿Αχαιοί is rare are akin to those discussed at the beginning of the preceding paragraph, and I therefore have classified ϝελικῶπες ᾿Αχαιοί as an accidental infrequent formula. The phrase ἥρωες ᾿Αχαιοί may be rarely needed because of the uncertainty over whether ϝελικῶπες ᾿Αχαιοί begins with a vowel or a consonant as the digamma begins to go; it only

[42] The Κρ- in Κρόνου πάις ἀγκυλομήτεω (which falls in the same colon as πατὴρ ἀνδρῶν τε Θεῶν τε) never makes position.

occurs once (and once in the *Odyssey*, not included in the count since ʼΑχαιοί in the *Odyssey* lacks a regular formula). When inflected the phrase is a regular formula in the accusative; but in the nominative there are only six occurrences all told of formulae that fill position 8-12. I classify μελιηδέα οἶνον without hesitation as an accidental infrequent formula, since it occurs 5 times; I am not sure why εὐήνορα οἶνον is needed rarely, but I infer that it is from the fact that it occurs only once.

There are 40 *semantic-aesthetic* alternatives to regular formulae, falling into two groups: formulae with epithets specific to the context (such as Θέτις κατὰ δάκρυ χέουσα, used instead of θεὰ Θέτις ἀργυρόπεζα) and formulae used for special effect (such as μεγάθυμος ʼΑχιλλεύς for πόδας ὠκὺς ʼΑχιλλεύς, mentioned above). We have 28 cases of the former, all of them used to say something particular to a situation that does not often arise, such as when Thetis is weeping. They may replace a regular formula in the same colon, as when Telemachos might have completed *Od.* 3.98 with πολύτλας δῖος ʼΟδύσσευς, but chose the much more appropriate πατὴρ ἐμὸς ἐσθλὸς ʼΟδυσσεύς. Or they may occupy an alternate colon, as when δεύτερος αὖτʼ ʼΑίας begins the line. Many combine two nouns, not in order to bring two separate ideas into a doubling formula, but to produce a larger single idea, as when Πρίαμος Πριάμοιό τε παῖδες is used to mean "Priam's family." I shall not list the rest; these examples should make clear what they are like.

There are just 12 cases where it seems appropriate speak of special effect.[43] Often, as with μεγάθυμος ʼΑχιλλεύς, the effect is merely the avoidance of ugliness. Parry calls attention to the simile where Zeus is said to move "the thick cloud from the high peak of a great mountain" (16.297-98). It would be unsuitable to call Zeus "cloud-gathering" here, but that is the regular formula for the colon with which the poet is confronted; and so we have instead "lightning-gathering,"

---

[43] The remarks in the next two paragraphs owe a great deal to the careful criticism of Richard Janko, who calls my attention to the large number of apparent *equivalent* formulae that are used for a special effect or are specific to the context. See Janko 1981 and Janko 1992:434, s.v. "equivalent formulae." To the extent, of course, that they are specific to a context or create a special effect they are not really equivalent.

στεροπηγερέτα.[44] In *Iliad* 2.645 Homer, had he used his regular formula, would have found himself saying, "Of the Cretans, Idomeneus, leader of the Cretans, was the leader." This would hardly have done, and so he dug into his bag of generic epithets and said instead, "Of the Cretans, Idomeneus the spear-famed was the leader."[45] The 2 formulae ἤνοπι χαλκῶι in both poems are used in unusual circumstances to avoid the military or death-dealing connotations of the regular formula. Then there are 4 formulae, employed just once, where the poet is not so much avoiding ugliness as using a colorful and unusual epithet: χρυσάμπυκας ἵππους (used of divine horses), Ϝερυσάρματας ἵππους (seemingly to bind two passages together),[46] ὑψηχέας ἵππους (a strange epithet, perhaps used in *Iliad* 5.772 to mark the divinity of the horses), and ὑψηχέες ἵπποι (in *Iliad* 23.27, perhaps marking the extraordary presence of the horses next to the pyre). The phrase ἀσπίδα θοῦριν, used just twice, has an epithet strange for the object, and we would therefore not expect to find it used often; in πίονα ἔργα the epithet gives the noun a sense unusual for it in the *Iliad*; in μέρμερα ἔργα the epithet itself is relatively rare.

There are 3 infrequent formulae whose existence I find it hard to account for: δῖοι 'Αχαιοί does not appear to be a necessary alternative to κοῦροι 'Αχαιῶν, nor Ζεὺς τερπικέραυνος (in both poems) to νεφεληγερέτα Ζεύς. In all 3 cases the meaning is different from the meaning of the regular formulae, but in the first there is a net loss of color, and in the second and third I cannot hear any gain. It is just possible that the force of n- as a double consonant in νεφεληγερέτα Ζεύς was being lost, but this is just guessing. To have three cases where we are just puzzled does not seem demoralizingly high.

---

[44] Parry 1971:187. Parry thought that Homer—or rather the tradition—was avoiding the doubled *sound* here, νεφέλην νεφελη-, and that may be the reason (see my next example); but what prevented him from considering the reason I prefer is his theory that the fixed epithet was not heard by the audience, and this view I find unacceptable: see Sale 1989:388-90 and Janko 1992:356.

[45] 'Ιδομενεὺς δουρικλυτός running from 3-8 actually occurs 5 times in all, once a few lines later in Book 2 and for the same reason, the other times in individual battle scenes where the poet, now fully equipped with the alternative to the infrequent formulae, apparently wished to use it when Idomeneus was fighting; the regular formula occurs off the battlefield and mostly to introduce speeches.

[46] *Iliad* 15.354 and 16.370, where the Trojans cross and recross the ditch; the epithet is used in the latter to extend the regular formula. See Janko 1992:266.

The other 42 infrequent formulae (7% of the total 652 for our 70 nouns) are probably infrequent formulae by accident. They fall in common major cola for their nouns, almost all put the noun at the localization-point, they are noun-epithets and their epithets are context-free; several are regular formulae in the other poem. These do not form exceptions to the general rule that infrequent formulae answer to rare needs; it is perfectly natural that a certain percentage of rare needs should arise by accident—should arise because the poet is using a given noun less often, or in different contexts, or in different metrical circumstances, than he might otherwise be doing.

It may be useful to examine a typical formulaic set. Ajax (the word Αἴας), for example, has 23 different formulae to go with his 80 total occurrences in the nominative. Not surprisingly, his localization is low, at 40%: because the word Αἴας can wander into 6 different parts of the line, it is free to develop infrequent formulae in 5 of them, more infrequent formulae than Ἀχιλλεύς, which occurs over twice as often (171 total occurences) but has a localization of 94%. Twenty-two of the 23 different Ajax formulae are infrequent; the one frequent formula, Τελαμώνιος Αἴας, sometimes extended with μέγας, occurs 21 times, filling the verse from the hephthemimeral caesura (or the trochaic caesura when extended) to the end.[47] His lack of a regular formula in 9-12 is compensated for by the infrequent formulae φαίδιμος Αἴας, which occurs 5 times. This accidental infrequent formulae would probably have been a regular formulae if Ajax' localization had not been so low, and the number of verse-positions he can occupy so large.

Ajax has 14 formulae that occur only once, 4 that occur twice, 2 that occur thrice, one 4 times, one 5; a total of 22 different infrequent formulae, 37 infrequent formulae-occurrences, somewhat lower than the average ratio of 1.96. Only 6 of the infrequent formulae fall in a major-colon, 27%, whereas 42% of the total of 652 infrequent formulae fall there. This low figure is largely accounted for by Ajax' low localization and the number of different positions he occupies. Two of the 6 major-colon formulae, 33%, are noun-verb formulae, as opposed to 56% for all 70 nouns; the numbers are too low for statistical significance. Not that Ajax lacks noun-verb formulae, quite the contrary; he has 11, but only 2 fall in major cola.

1. Just 3 (21%) of the 14 once-only formulae fill a major colon; in

---

[47] It is important to keep in mind that the extension of a formula, regular or infrequent, is not counted as a different formula, since it contains a formula that is exactly repeated. See above, note 2, and Sale 1989:382.

contrast, 30% of the total 310 once-only do this. None occupies the same colon as any other: we have 14 formulae in 14 different positions. Two (67%) of the major-colon are noun-verb formulae; contrast 59% of the total 94 major-colon once-only formulae. The other is δεύτερος αὖτ' Αἴας, specific to the context; it overlaps Αἴας διογενής metrically but not, of course, semantically. There is no violation of the principle of economy. Two put the noun at the localization-point; in the others Αἴας has wandered to 3 different unusual positions.

2. Two (50%) of the 4 twice-only formulae fill a major colon, and 2 do not; in contrast, 44% of the total 190 twice-only formulae do this. Again we find none in the same colon as any other: 4 formulae, 4 different positions. Neither of the major-colon formulae is a noun-verb formula; contrast 56%. One is Αἴας διογενῆς, a major-colon noun-epithet formula occupying a rare colon; the other is Τελαμώνιος ἄλκιμος Αἴας. This can be analyzed as the regular formula in a new position, after the penthemimeral, and separated, in which case it occupies a rare major colon; or it can be seen as ἄλκιμος Αἴας extended, occupying a common major colon and consisting of an accidental infrequent formulae. One of the twice-only formulae puts the noun at the localization-point; in the others Αἴας has wandered to 2 different unusual positions.

3. Neither of the two three-times formulae fills a major colon; contrast 50% of the total of 76 thrice-only. Neither is in the same colon as the other. Neither puts the noun at the localization-point; but both put it in the same unusual position.

4. The four-times formula does not fall in a major colon; contrast 70%. It is not a noun-verb formula; in fact it is Αἴας. . .Τελαμώνιος, the regular formula in a new position, 1-8, and separated. The noun is not at the localization-point.

5. The five-times formula falls in a major colon; so do 80% of the total of 30 five-times. It is not a noun-verb; contrast 50% of the total 24 major-colon formulae. It is φαίδιμος Αἴας, which we consider an accidental infrequent formulae. The noun is at the localization-point.